

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 13:53:16 ; Search time 4521.53 Seconds  
(without alignments)  
16.136 Million cell updates/sec

Title: US-09-164-714-13

Perfect score: 18  
Sequence: 1 ggtgtgtgtgtgtgtt 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 119968

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST: \*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_est48:\*  
83: gb\_est49:\*  
84: gb\_est50:\*  
85: gb\_est51:\*  
86: em\_est34:\*  
87: em\_est35:\*  
88: em\_est36:\*  
89: em\_est37:\*  
90: gb\_est52:\*  
91: gb\_est53:\*  
92: gb\_est54:\*  
93: gb\_est55:\*  
94: gb\_est56:\*  
95: em\_est38:\*  
96: em\_est39:\*  
97: em\_est40:\*  
98: em\_est41:\*  
99: em\_est42:\*  
100: em\_est43:\*  
101: em\_est44:\*  
102: gb\_est57:\*  
103: gb\_est58:\*  
104: em\_est39:\*  
105: gb\_est59:\*  
106: gb\_est60:\*  
107: gb\_est61:\*  
108: gb\_est62:\*  
109: gb\_est63:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

Result

Query

SUMMARIES

No.	Score	Match	Length	DB	ID	Description
C 1	15.4	85.6	77	40	D43091	D43091 D43091 Rice
C 2	14.8	82.2	64	50	F30799	F30799 HSPD21447 H
C 3	14	77.8	94	49	F24375	F24375 HSPD10647 H
C 4	13.8	76.7	58	45	A1343303	A1343303 tbr2909.x
C 5	13.8	76.7	82	38	AA748306	AA748306 oas4a10.s
C 6	13.4	74.4	46	34	AA511536	AA511536 vj24e05.f
C 7	13.4	74.4	74	63	A1971085	A1971085 wr3906.x
C 8	13.4	74.4	77	20	D20607	D20607 HUMG501582
C 9	13.4	74.4	79	29	AA160831	AA160831 z662f02.s
C 10	13.4	74.4	89	44	A1320114	A1320114 b9g05mm.f
C 11	13.2	73.3	76	60	A1789049	A1789049 uk50d02.y
C 12	13.2	73.3	76	62	A1924620	A1924620 wn57d08.x
C 13	13.2	73.3	95	37	AA680947	AA680947 lmfFrm055
C 14	13.2	73.3	97	37	AA680946	AA680946 lmfFrm055
C 15	13.2	73.3	100	50	A1682938	A1682938 tw47d01.x
C 16	13	72.2	25	40	A1000095	A1000095 os61a08.s
C 17	13	72.2	74	37	AA680772	AA680772 lmfFrm030
C 18	13	72.2	99	35	AA591720	AA591720 v148d01.r
C 19	12.8	71.1	46	63	A1957210	A1957210 ul17a04.x
C 20	12.8	71.1	87	94	A0072951	A0072951 EP(2)2165
C 21	12.4	68.9	52	21	T92783	T92783 ye27b02.s1
C 22	12.4	68.9	70	62	A1892436	A1892436 mm99e04.y
C 23	12.4	68.9	76	21	R15882	R15882 ya47a12.r1
C 24	12.4	68.9	87	43	A1161047	A1161047 qc93d02.x
C 25	12.4	68.9	91	40	AA948011	AA948011 oq58906.s
C 26	12.2	67.8	28	48	A1590364	A1590364 ts09e12.x
C 27	12.2	67.8	34	44	A1287522	A1287522 qv86g04.x
C 28	12.2	67.8	37	41	A1018534	A1018534 ou24h12.x
C 29	12.2	67.8	37	47	A1475986	A1475986 t196c09.x
C 30	12.2	67.8	46	23	H24541	H24541 y130f06.r1
C 31	12.2	67.8	46	41	A1033350	A1033350 ox02e11.s
C 32	12.2	67.8	51	79	AA232755	AA232755 f130a01.x
C 33	12.2	67.8	52	23	H24483	H24483 y130a09.r1
C 34	12.2	67.8	60	24	D69796	D69796 CELK092C6F
C 35	12.2	67.8	73	49	A1635197	A1635197 tz22d03.x
C 36	12.2	67.8	74	40	AA984916	AA984916 am63a09.s
C 37	12.2	67.8	76	26	W98562	W98562 mg14f02.r1
C 38	12.2	67.8	76	43	A1194156	A1194156 ue78f02.r
C 39	12.2	67.8	80	44	A1267776	A1267776 aq34a07.x
C 40	12.2	67.8	80	94	AA0073477	AA0073477 EP(2)2306
C 41	12.2	67.8	80	94	AA0073491	AA0073491 EP(2)2322
C 42	12.2	67.8	80	94	AA0073569	AA0073569 EP(2)2416
C 43	12.2	67.8	87	92	AQ911807	AQ911807 LMAJFV1.1
C 44	12.2	67.8	88	45	A1354446	A1354446 qu17f10.x
C 45	12.2	67.8	90	22	R40179	R40179 yf70e09.s1

## ALIGNMENTS

```

RESULT 1
LOCUS D43091 77 bp mRNA EST 04-MAY-1998
DEFINITION D43091 Rice callus cDNA (H.Uchimiyu) Oryza sativa cDNA clone AD544,
mRNA sequence.
ACCESSION D43091
VERSION D43091.1 GI:3107351
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
REFERENCE 1 (bases 1 to 77)
AUTHORS Uchimiyu, H.
TITLE On nucleotide sequence of Oryza sativa
JOURNAL Unpublished (1994)
COMMENT On Jan 14, 1998 this sequence version replaced gi:2754210.
Contact: Hirofumi Uchimiyu
Institute of Mol. & Cell. Bioscience, Department of Cellular
Function

```

The University of Tokyo  
1-1-1 Yayoi, Bunkyo-ku, Tokyo 113, Japan  
Tel: 03-3812-2111 (ex.7844)  
Fax: 03-3812-2910  
Email: huchimiyu@tansei.cc.u-tokyo.ac.jp  
PROJECT= Uchimiyu.

FEATURES  
source  
1..77  
Location/Qualifiers

/organism="Oryza sativa"  
/db\_xref="taxon:4530"  
/clone="AD544"  
/clone\_lib="Rice callus cDNA (H.Uchimiyu)"  
/tissue\_type="callus"

BASE COUNT 22 a 25 c 9 g 14 t 7 others

ORIGIN

Query Match 85.6%; Score 15.4; DB 40; Length 77;  
Best Local Similarity 94.1%; Pred. No. 2.2e+03;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ggtgtgtgtgtgtgtt 17  
|||||  
Db 61 GGTGTGTGTGTGTGT 45

RESULT 2

LOCUS F30799 64 bp mRNA EST 13-MAY-1999  
DEFINITION HSPD21447 HM3 Homo sapiens cDNA clone s4000088F12, mRNA sequence.  
ACCESSION F30799  
VERSION F30799.1 GI:4816425  
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 64)

AUTHORS Lanfranchi, G., Muraro, F., Caldara, F., Pacchioni, B., Pallavicini, A.,  
Pandolfo, D., Toppo, S., Trevisan, S., Scarso, S. and Valle, G.  
IDENTIFICATION OF 4370 EXPRESSED SEQUENCE TAGS FROM A  
3'-END-SPECIFIC cDNA LIBRARY OF HUMAN SKELETAL MUSCLE BY DNA  
SEQUENCING AND FILTER HYBRIDIZATION  
Genome Res. 6 (1), 35-42 (1996)

JOURNAL  
MEDLINE  
96276048  
ON MAY 18, 1998 THIS SEQUENCE VERSION REPLACED GI:3137194.

COMMENT

Contact: Valle G.

CIRIBI Biotechnology Centre

University of Padua

Via Trieste 75, 35121 Padua, Italy

ABI Chromatograms and other information are available on WWW at

http://group.bio.unipad.it.

FEATURES

source

Location/Qualifiers

1..64

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="s4000088F12"

/clone\_lib="HM3"

/sex="female"

/tissue\_type="pectoral muscle (after mastectomy)"

/note="vector: pcDNA1 (Invitrogen); Site\_1: BstXI;

Site\_2: NotI; The library was not constructed nor normalized.

Lanfranchi. This library is not constructed nor normalized.

The first strand cDNA was primed with a biotinylated

oligo-dT-NotI primer

(5'-biotin-AACCGGCTGAGCGCCGCTTTTCTTTTCTTTT-3'). The

ds cDNA was sonicated and size-selected in the range

350-550 bp. The 3' specific fragments were selected by

streptavidin coated magnetic beads, ligated to

non-palindromic BstXI adapters, NotI digested and

directionally cloned into BstXI-NotI cut pcDNA1 vector."

BASE COUNT 7 a 15 c 24 g 18 t

ORIGIN

Query Match 82.2%; Score 14.8; DB 50; Length 64;  
Best Local Similarity 88.9%; Pred. No. 3.9e+03;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ggtgtgtgtgtgtgtgtt 18  
||||| |||||||  
Db 10 GGTGTGTGAAGTTGGCTTT 27

RESULT 3  
F24375 94 bp mRNA EST 13-MAY-1999  
LOCUS F24375  
DEFINITION HSPD10647 HM3 Homo sapiens cDNA clone s4000008G11, mRNA sequence.  
ACCESSION F24375  
VERSION F24375.1 GI:4810001  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 94)  
Lanfranchi, G., Muraro, T., Caldara, F., Pacchioni, B., Pallavicini, A.,  
Pandolfo, D., Toppo, S., Trevisan, S., Scarso, S. and Valle, G.  
Identification of 4370 expressed sequence tags from a  
3'-end-specific cDNA library of human skeletal muscle by DNA  
sequencing and filter hybridization  
Genome Res. 6 (1), 35-42 (1996)  
96276048  
JOURNAL MEDLINE  
COMMENT On Feb 18, 1999 this sequence version replaced gi:4314769.  
Contact: Valle G.  
CIRI Biotechnology Centre  
University of Padua  
Via Trieste 75, 35121 Padua, Italy  
ABI Chromatograms and other information are available on WWW at  
http://grup.bio.unipd.it  
Location/Qualifiers  
FEATURES  
source  
1. 94  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="s4000008G11"  
/clone\_lib="HM3"  
/sex="female"  
/tissue\_type="pectoral muscle (after mastectomy)"  
/note="Vector: pCDNAII (Invitrogen); Site 1: BstXI;  
Site 2: NotI; The library was constructed by G.  
Lanfranchi. This library is not subcloned nor normalized.  
The first strand cDNA was primed with a biotinylated  
oligo-dT-NotI primer  
(5'-biotin-AACCGGCTCGAGCGCGCTTTT-3'). The  
ds cDNA was sonicated and size-selected in the range  
350-550 bp. The 3' specific fragments were selected by  
streptavidin coated magnetic beads, ligated to  
non-palindromic BstXI adapters, NotI digested and  
directionally cloned into BstXI-NotI cut pCDNAII vector."

BASE COUNT 27 a 28 c 29 g 10 t

Query Match 77.8%; Score 14; DB 49; Length 94;  
Best Local Similarity 100.0%; Pred. No. 9e+03;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 ggtgtgtgtgtgtgtt 17  
||||| |||||||  
Db 14 GTTGTGTGTGCTT 1

RESULT 4  
AI343303 58 bp mRNA EST 08-APR-1999  
LOCUS AI343303/c  
DEFINITION tb92909.x1 NCI\_CGAP\_Lu25 Homo sapiens cDNA clone IMAGE:2061856 3'

similar to TR:Q33578 Q33578 KINETOPLAST CR5 ;, mRNA sequence.

ACCESSION AI343303  
VERSION AI343303.1 GI:4080509  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 58)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL MEDLINE  
COMMENT On Jan 17, 1998 this sequence version replaced gi:2150355.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: I.M.A.G.E. Consortium, LNL.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www.bio.lnl.gov/bdrp/image/image.html

Trace considered overall poor quality  
Seq primer: -400P from Gibco  
High quality sequence stop: 1.  
Location/Qualifiers  
FEATURES  
source  
1. 58  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2061856"  
/clone\_lib="NCI\_CGAP\_Lu25"  
/tissue\_type="bronchioalveolar carcinoma"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: PAMP1; mRNA made from lung  
carcinoma tissue, cDNA made by oligo-dT priming.  
Directionally cloned. Size-selected on agarose gel,  
average insert size 500 bp. Primary library,  
non-amplified."

BASE COUNT 34 a 15 c 8 g 1 t

Query Match 76.7%; Score 13.8; DB 45; Length 58;  
Best Local Similarity 88.2%; Pred. No. 1e+04;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 ggtgtgtgtgtgtt 18  
||||| |||||||  
Db 31 GTTGTGTGTGTATTT 15

RESULT 5  
AA748306 82 bp mRNA EST 23-JAN-1998  
LOCUS AA748306/c  
DEFINITION ca54a10.s1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:1308762 3'  
similar to WP:C36B1.12 CE05379 ;, mRNA sequence.  
ACCESSION AA748306  
VERSION AA748306.1 GI:2788264  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 82)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL MEDLINE  
COMMENT On Jan 14, 1998 this sequence version replaced gi:1798074.  
Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/db/brp/image/image.html](http://www-bio.llnl.gov/db/brp/image/image.html)

Trace considered overall poor quality  
 Insert Length: 438 Std Error: 0.00  
 Seq primer: -40m13 fwd. Et from Amersham  
 High quality sequence stop: 1.

## FEATURES

source

location/qualifiers

```
1. 82
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1308762"
/clone_lib="NCI CGAP GCBI"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pRT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI). Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTCACATCTGAAAGTGAGCGCGCCGTCATTTTCTTTTCTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pRT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
```

## BASE COUNT

49 a 22 c 10 g 1 t

## ORIGIN

1 t

Query Match 76.7%; Score 13.8; DB 38; Length 82;  
 Best Local Similarity 88.2%; Pred. No. 1.1e+04;  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 ggtgtgtgtgtgtgtt 17  
 ||||| ||||| |||||  
 Db 23 GGTGTGTGTGTGCTT 7

RESULT 6 46 bp mRNA EST 08-JUL-1997  
 AA511536  
 LOCUS vj24e05.r1 Soares mouse NBMH Mus musculus cDNA clone IMAGE:922688  
 DEFINITION 5' similar to SW:NM\_002008 P03893 NMDH-UBIQUITONE OXIDOREDUCTASE  
 CHAIN 2 ; mRNA sequence.

ACCESSION AA511536  
 VERSION AA511536.1 GI:2249390  
 KEYWORDS EST.

## ORGANISM

house mouse.  
 Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 46)  
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubugue, T.,  
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.

## AUTHORS

The WashU-HHMT Mouse EST Project

TITLE Unpublished (1996)  
 JOURNAL Contact: Maria M/Mouse EST Project  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1395342.

WashU-HHMT Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium ([Info@image.llnl.gov](mailto:Info@image.llnl.gov)) for further information.  
 MGI:534904

Trace considered overall poor quality  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -28m13 rev2 Et from Amersham  
 High quality sequence stop: 1.

## FEATURES

source

location/qualifiers

```
1. 46
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:922688"
/clone_lib="Soares mouse NBMH"
/sex="male"
/tissue_type="heart"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pRT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTCACATCTGAAAGTGAGCGCGCCGTCATTTTCTTTTCTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pRT73 vector. RNA
provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."
```

## BASE COUNT

15 a 16 c 5 g 10 t

## ORIGIN

10 t

Query Match 74.4%; Score 13.4; DB 34; Length 46;  
 Best Local Similarity 93.3%; Pred. No. 1.5e+04;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 4 gttgtgtgtgtgtt 18  
 ||||| ||||| |||||  
 Db 43 GTTGTGTGTGCTAT 29

## RESULT 7

AI971085 74 bp mRNA EST 25-AUG-1999  
 AI971085  
 LOCUS w23906.x1 NCI CGAP Pr28 Homo sapiens cDNA clone IMAGE:2488570 3'

DEFINITION similar to gb:M62402 INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 6  
 PRECURSOR (HUMAN); mRNA sequence.

## ACCESSION

AI971085  
 VERSION AI971085.1 GI:5767911  
 KEYWORDS EST.

## ORGANISM

human.  
 Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 74)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index

## AUTHORS

Unpublished (1997)  
 On Dec 20, 1995 this sequence version replaced gi:1130903.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550

Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.



DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution Information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbcp/image/image.html](http://www.bio.llnl.gov/dbcp/image/image.html)

Trace considered overall poor quality  
Seq primer: -40UP from Gibco  
High quality sequence stop: 1.

## FEATURES

source

1. 74

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:248570"  
/clone\_lib="NCI\_CGAP\_Pr28"  
/sex="male"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)  
with a modified polylinker; Plasmid DNA from the  
normalized library NCI\_CGAP\_Pr22 was prepared, and ss  
circles were made in vitro. Following HAP purification,  
this DNA was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from a pool  
of 5,000 clones made from the same library (cloneids  
985608-986759, 1101192-1101959, and 1217928-1220615).  
Subtraction by Bento Soares and M. Fatima Bonaldo."

## BASE COUNT

20 a 23 c 19 g 12 t

ORIGIN

## Query Match

Best Local Similarity 74.4%; Score 13.4; DB 63; Length 74;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

## Qy 2 ggtgtgtgtgtgct 16

Db 38 GTCTGTGTGTGGCT 24

## RESULT 8

LOCUS D20607/c

DEFINITION D20607 77 bp mRNA EST 30-JUL-1996  
HUMS01582 Human promyelocyte Homo sapiens cDNA clone pm152 3',  
mRNA sequence.

ACCESSION D20607

VERSION D20607.1 GI:501703

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 77)  
Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y.,  
Yoshinari,H., Arimoto,J. and Matsubara,K.  
Gene expression of human promyelocytic cell line HL60 before and  
after induction of differentiation. A new application of 3'directed  
cDNA sequencing  
Unpublished (1993)  
Contact: Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y.,  
Yoshinari,H., Arimoto,J. and Matsubara,K.  
Institute for Molecular and Cellular Biology  
Osaka University  
3-1 Yamada-oka,Suita,Osaka 565,Japan.

FEATURES

source

1. 77

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="pm152"  
/clone\_lib="Human promyelocyte"  
/note="Female, adult, cell\_line = HL60, cell\_type =  
promyelocyte."

BASE COUNT 27 a 22 c 10 g 17 t 1 others

ORIGIN

Query Match 74.4%; Score 13.4; DB 20; Length 77;  
Best Local Similarity 87.5%; Pred. No. 1.6e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

## Qy 1 ggtgtgtgtgtgct 16

Db 41 GGNCTGTGTGTGGCT 26

## RESULT 9

LOCUS AA160831

DEFINITION AA160831 79 bp mRNA EST 16-DEC-1996  
Z062f02.s1 Stratiagene pancreas (#937208) Homo sapiens cDNA clone  
IMAGE:591483 3' similar to SW:RS5\_HUMAN P46782 40S RIBOSOMAL  
PROTEIN S5.; mRNA sequence.

ACCESSION AA160831

VERSION AA160831.1 GI:1736267

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

MEDLINE

GENOME

RES.

6 (9), 807-828 (1996)

97044478

On May 18, 1995 this sequence version replaced gi:811261.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: estewatson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium ([infoimage.llnl.gov](http://infoimage.llnl.gov)) for further information.

Trace considered overall poor quality

Possible reversed clone: similarity on wrong strand

Seq primer: -40M13 fwd. from Amersham

High quality sequence stop: 1.

Location/Qualifiers

1. 79

/organism="Homo sapiens"  
/db\_xref="GDB:4622869"  
/db\_xref="taxon:9606"  
/clone="IMAGE:591483"  
/clone\_lib="Stratiagene pancreas (#937208)"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: pancreas; Vector: pBluescript SK-; Site: 1:  
ECORI; Site: 2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT. Pancreatic adenocarcinoma cell line. Average  
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor  
sequence: 5' GAATTCGGCAGCG 3' -3' adaptor sequence: 5'  
CTCGAGTCTTTTCTTTTCTTTTCTTTT 3'."

## BASE COUNT

16 a 25 c 21 g 13 t 4 others

ORIGIN

## Query Match

Best Local Similarity 74.4%; Score 13.4; DB 29; Length 79;

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

## Qy 3 tgtgtgtgtgtgctt 18

Db 49 TGTGTGTGTGTGGCT 64

RESULT	10	AI320114	89 bp	mRNA	EST	18-DEC-1998
LOCUS	AI320114/c					
DEFINITION	b905nm.fl	Neurospora crassa morning	cdna library	Neurospora crassa		
ACCESSION	AI320114					
VERSION	AI320114.1	GI:4036096				
KEYWORDS	EST.					
SOURCE	Neurospora crassa.					
ORGANISM	Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae;					
REFERENCE	Neurospora.					
AUTHORS	1 (bases 1 to 89)					
TITLE	Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C. and Roe,B.A.					
JOURNAL	Two Neurospora crassa EST Databases					
COMMENT	Unpublished (1998)					
	Contact: Bruce A. Roe, University of Oklahoma, broeeu.edu					
	Department of Chemistry and Biochemistry					
	Advanced Center for Genome Technology, University of Oklahoma					
	620 Parrington Oval, Norman, OK 73019, USA					
	Tel: 405 325 4912					
	Fax: 405 325 7762					
	Email: broeeu.edu					
	We anticipate the future release of the cdna clones to the Fungal					
	Genetics Stock Center					
	Seq primer: Universal Forward Primer					
	High quality sequence stop: 78.					
FEATURES	Location/Qualifiers					
SOURCE	1..89					
	/organism="Neurospora crassa"					
	/strain="pd, Irq? A"					
	/db_xref="taxon:5141"					
	/clone="b905nm"					
	/clone_1pb="Neurospora crassa morning cdna library"					
	/tissue_type="tissue harvested following 22hr growth in					
	dark"					
	/note="Vector: pBluescript SK-; Site_1: XbaI; Site_2:					
	EcoRI; See: Bell-Pedersen,D., et al. PNAS 93:13096,1996.					
	5' end of cdna cloned into EcoRI site of pBluescript; 3'					
	end of cdna cloned into EcoRI site of pBluescript"					
BASE COUNT	20 a 35 c 15 g 19 t					
ORIGIN						
Query Match	74.4%;	Score 13.4;	DB 44;	Length 89;		
Best Local Similarity	93.3%;	Pred. No. 1.6e+04;				
Matches 14;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;		
Oy	1 ggtgtgtgtgtgc 15					
Db	21 ggtgtgtgtgtgc 7					
RESULT 11						
LOCUS	AI1789049/c					
DEFINITION	AI1789049	76 bp	mRNA	EST	02-JUL-1999	
	U50d02.y1	Sugano mouse kidney mKia	Mus musculus	cdna clone		
	IMAGE:1972419 5'	similar to TR:Q28170	Q28170	POLYUBIDUITIN ;		mRNA
ACCESSION	AI1789049					
VERSION	AI1789049					
KEYWORDS	EST.					
SOURCE	house mouse.					
ORGANISM	Mus musculus					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;					
	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
	1 (bases 1 to 76)					
	Marras,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylje,T.,					
	Underwood,K., Stepcoe,M., Theisling,B., Allen,M., Bowers,Y.,					
	Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R.,					
	Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,					

TITLE	Waterston, R. and Wilson, R									
JOURNAL	The WashU-NCI Mouse EST Project 1999									
COMMENT	Unpublished (1999) On Dec 20, 1995 this sequence version replaced gi:1135238. Contact: Marra M/WashU-NCI Mouse EST Project 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: mousest@waterston.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. MGI:989159									
FEATURES	Trace considered overall poor quality possible reversed clone; similarity on wrong strand Seq primer: custom primer used High quality sequence stop: 1. Location/Qualifiers 1..76									
SOURCE	/organism="Mus musculus" /strain="C57BL" /db_xref="taxon:10090" /clone="IMAGE:1972419" /clone_lib="Sugano mouse kidney mk1a" /sex="female" /dev_stage="adult" /lab_host="DH10B" /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII (CAGCTGTG); Site_2: DraIII (CACCAGTGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CAGCTGTG, 3' site CACCAGTGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' and primer CTCTGCTCTCTAAAGCTGCG and 3' end primer CGACCTCGAGCTCGAGCACA."									
BASE COUNT	17 a 19 c 19 g 21 t									
ORIGIN										
Query Match	73.3%; Score 13.2; DB 60; Length 76;									
Best Local Similarity	83.3%; Pred. No. 1.9e+04;									
Matches 15; Conservative	0; Mismatches 3; Indels 0; Gaps 0									
QY	1 ggtgtgtgtgttgccttc 18 									
Db	73 GGTGTGTGTGTGCTGCTGT 56									
RESULT 12										
LOCUS	AI924620/c 76 bp mRNA EST 02-SEP-1999									
DEFINITION	wn57gd08.x1 NCI_CGAP Lu19 Homo sapiens cDNA clone IMAGE:2449551 3'									
ACCESSION	AI924620									
VERSION	AI924620.1 GI:5660584									
KEYWORDS	EST.									
SOURCE	human.									
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Euthetia; Primates; Catarrhini; Homiidae; Homo.									
REFERENCE	1 (bases 1 to 76) NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .									
ADTHOS	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),									
TITLE	Unpublished (1997)									
JOURNAL	On Dec 20, 1995 this sequence version replaced gi:1135542.									
COMMENT	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550									

Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL at:  
www.bio.lnl.gov/dbp/image/image.html

Trace considered overall poor quality  
Seq primer: -40UP from Gibco  
High quality sequence stop: 1.

## FEATURES

## source

1. /76  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2449551"  
/clone\_lib="NCI CGAP\_Lu19"  
/tissue\_type="squamous cell carcinoma, poorly  
differentiated (4 pooled tumors, including primary and  
metastatic)"  
/dev\_stage="adult"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pF773D-Pac (Pharmacia) with a  
modified polylinker; 1st strand cDNA was prepared from a  
pooled lung tumor tissue, and was then primed with a Not I  
- oligo(dT) primer. Double-stranded cDNA was ligated to  
Eco RI adaptors (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of the modified  
pF773 vector. Library went through one round of  
normalization. Library constructed by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT 28 a 24 c 16 g 8 t  
ORIGIN

Query Match 73.3%; Score 13.2; DB 62; Length 76;  
Best Local Similarity 83.3%; Pred. No. 1.9e+04;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ggtgtgtgtgtgtgttt 18  
||||| ||| ||| |||  
Db 65 GGTGTGTCTCTGCTTT 48

RESULT 13  
AA680947/c

LOCUS AA680947 95 bp mRNA EST 09-DEC-1998  
DEFINITION LmfrAm0555 Leishmania major Amastigote Lambda Zap II library  
Leishmania major cDNA clone R69 5', mRNA sequence.

ACCESSION AA680947  
VERSION AA680947.1 GI:2662952  
KEYWORDS EST.  
SOURCE Leishmania major.  
ORGANISM Leishmania major  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Leishmania.

REFERENCE 1 (bases 1 to 95)  
AUTHORS Norrish,A.R., Dyall,S.D., Smith,D.F. and Blackwell,J.M.  
TITLE Analysis of Leishmania Major Amastigote Expressed Sequence Tags  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1997 this sequence version replaced g1:2264540.  
Contact: Blackwell JM  
Wellcome Trust/MRC Building, Addenbrooke's Hospital, Hills Road,  
Cambridge CB2 2XY, UK  
Tel: 01223 336 143  
Fax: 01223 331 206  
Email: jmb37@cus.cam.ac.uk  
PCR Primers  
FORWARD: GTAAACGACGACGACG  
BACKWARD: GGAACACGTATGACCATG

Seq primer: AATTACCTTCACATAAGG  
High quality sequence stop: 95.

## FEATURES

## source

1. /95  
/organism="Leishmania major"  
/strain="Friedlin"  
/db\_xref="taxon:5664"  
/clone="R69"  
/clone\_lib="Leishmania major Amastigote Lambda Zap II  
library"  
/cell\_type="Amastigote"  
/note="Vector: Lambda Zap II; Site\_1: XhoI; Site\_2: NotI"

BASE COUNT 54 a 39 c 2 g 0 t  
ORIGIN

Query Match 73.3%; Score 13.2; DB 37; Length 95;  
Best Local Similarity 83.3%; Pred. No. 2e+04;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ggtgtgtgtgtgtgttt 18  
||||| ||| ||| |||  
Db 53 GGTGTGTGTGTGTGT 36

RESULT 14  
AA680946/c

LOCUS AA680946 97 bp mRNA EST 09-DEC-1998  
DEFINITION LmfrAm0554 Leishmania major Amastigote Lambda Zap II library  
Leishmania major cDNA clone R68 5', mRNA sequence.

ACCESSION AA680946  
VERSION AA680946.1 GI:2662951  
KEYWORDS EST.  
SOURCE Leishmania major.  
ORGANISM Leishmania major  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Leishmania.

REFERENCE 1 (bases 1 to 97)  
AUTHORS Norrish,A.R., Dyall,S.D., Smith,D.F. and Blackwell,J.M.  
TITLE Analysis of Leishmania Major Amastigote Expressed Sequence Tags  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1997 this sequence version replaced g1:2264539.  
Contact: Blackwell JM  
Wellcome Trust/MRC Building, Addenbrooke's Hospital, Hills Road,  
Cambridge CB2 2XY, UK  
Tel: 01223 336 143  
Fax: 01223 331 206  
Email: jmb37@cus.cam.ac.uk  
PCR Primers  
FORWARD: GTAAACGACGACGACG  
BACKWARD: GGAACACGTATGACCATG  
Seq primer: AATTACCTTCACATAAGG  
High quality sequence stop: 97.

FEATURES  
source

## source

1. /97  
/organism="Leishmania major"  
/strain="Friedlin"  
/db\_xref="taxon:5664"  
/clone="R68"  
/clone\_lib="Leishmania major Amastigote Lambda Zap II  
library"  
/cell\_type="Amastigote"  
/note="Vector: Lambda Zap II; Site\_1: XhoI; Site\_2: NotI"

BASE COUNT 56 a 39 c 2 g 0 t  
ORIGIN

Query Match 73.3%; Score 13.2; DB 37; Length 97;  
Best Local Similarity 83.3%; Pred. No. 2e+04;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ggtgtgtgtgtgtgttt 18

Db 53 GGCTTGTGTGTGGTGT 36

## RESULT 15

LOCUS

AI682938 100 bp mRNA

EST 15-DEC-1999

DEFINITION

tw47d01.x1 NCI\_CGAP\_Ut1 Homo sapiens cDNA clone IMAGE:2262817 3' similar to TR:Q91236 Q91236 ZONA PELUCIDA PROTEIN ; contains MER22.b2 MSRI repetitive element ; , mRNA sequence.

ACCESSION

AI682938

VERSION

AI682938.1 GI:4893120

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE

1 (bases 1 to 100)

AUTHORS

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL

Tumor Gene Index

COMMENT

Unpublished (1997)

On May 18, 1998 this sequence version replaced gi:3137345.

Contact: Robert Strausberg, Ph.D.

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/BLNT at:

[www.bio.lnl.gov/bbrp/image/image.html](http://www.bio.lnl.gov/bbrp/image/image.html)

Trace considered overall poor quality

Insert Length: 1340 Std Error: 0.00

Seq Primer: -400P from Gibco

High quality sequence stop: 1.

## FEATURES

source

1..100

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2262817"

/clone\_lib="NCI\_CGAP\_Ut1"

/tissue\_type="well-differentiated endometrial

adenocarcinoma, 7 pooled tumors"

/lab\_host="DH10B"

/note="Organ: uterus; Vector: PCMV-SPORT6; Site\_1: SalI;

Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.75 kb. Life technologies catalog #:

11538-014"

BASE COUNT

33 a 47 c 20 g 0 t

ORIGIN

## Query Match

Best Local Similarity 73.3%; Score 13.2; DB 50; Length 100;

Matches 15; Conservative

0; Mismatches 3; Indels 0; Gaps 0;

QY 1

gggtgtgtgtgtgtgttt 18

Db 86

GGCTTGTGTGTGGTGT 69

Search completed: June 4, 2000, 13:53:19  
Job time: 20943 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:05:35 ; Search time 1236.38 Seconds  
(without alignments)  
-14.163 Million cell updates/sec

Title: US-09-164-714-13  
Perfect score: 18  
Sequence: 1 ggtgtgtgtgtgtgttt 18

Scoring table: IDENTITY-NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 356616

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_com:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pl4:\*  
11: gb\_pl5:\*  
12: gb\_pl6:\*  
13: gb\_pl7:\*  
14: gb\_pl8:\*  
15: gb\_pl9:\*  
16: gb\_pl10:\*  
17: gb\_pl11:\*  
18: gb\_pl12:\*  
19: gb\_pl13:\*  
20: gb\_pl14:\*  
21: gb\_pl15:\*  
22: gb\_pl16:\*  
23: gb\_pl17:\*  
24: gb\_pl18:\*  
25: gb\_pl19:\*  
26: gb\_pl20:\*  
27: gb\_pl21:\*  
28: gb\_pl22:\*  
29: gb\_pl23:\*  
30: gb\_pl24:\*  
31: gb\_pl25:\*  
32: gb\_pl26:\*  
33: gb\_pl27:\*  
34: gb\_pl28:\*  
35: gb\_pl29:\*  
36: gb\_pl30:\*  
37: gb\_pl31:\*  
38: gb\_pl32:\*  
39: gb\_pl33:\*  
40: gb\_pl34:\*  
41: gb\_pl35:\*  
42: gb\_pl36:\*  
43: gb\_pl37:\*  
44: gb\_pl38:\*

45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pl5:\*  
52: gb\_htg8:\*  
53: gb\_htg9:\*  
54: gb\_htg10:\*  
55: gb\_htg11:\*  
56: gb\_htg12:\*  
57: gb\_htg13:\*  
58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14.4	80.0	90	5	136132 Sequence 16
2	13.8	76.7	36	5	127773 Sequence 5
3	13.8	76.7	36	5	127775 Sequence 7
4	13.2	73.3	20	5	176185 Sequence 4
5	13.2	73.3	24	5	127784 Sequence 16
6	13.2	73.3	41	5	A51237 Sequence 8
7	13.2	73.3	45	5	A77178 Sequence 6
8	13.2	73.3	45	5	A77180 Sequence 8
9	13.2	73.3	45	5	120213 Sequence 6
10	13.2	73.3	45	5	120215 Sequence 8
11	13.2	73.3	51	5	A12323 oligonucleo
12	13.2	73.3	51	5	A12324 oligonucleo
13	13.2	73.3	51	5	A12326 oligonucleo
14	13.2	73.3	51	5	A12397 oligonucleo
15	13.2	73.3	92	12	D55656S10
16	12.8	71.1	26	5	AR018921 Sequence
17	12.8	71.1	26	5	AR066549 Sequence
18	12.8	71.1	26	5	I72266 Sequence 43
19	12.8	71.1	26	5	I76833 Sequence 43
20	12.8	71.1	66	5	A02001 Artificial
21	12.8	71.1	66	5	A06447 Artificial
22	12.8	71.1	66	5	AF027108 Ustilago
23	12.8	71.1	76	5	E02399 DNA encodin
24	12.8	71.1	77	5	E01808 DNA encodin
25	12.8	71.1	77	5	E02146 DNA encodin
26	12.8	71.1	77	5	E02398 DNA encodin
27	12.8	71.1	93	5	A22901 H. insoliens
28	12.8	71.1	93	5	E03016 DNA encodin
29	12.8	68.9	54	5	I50003 Sequence 1
30	12.4	68.9	62	5	I50012 Sequence 10
31	12.4	68.9	63	5	I50013 Sequence 11
32	12.4	68.9	85	5	AR051981 Sequence
33	12.4	68.9	85	5	I52155 Sequence 65
34	12.4	68.9	90	9	HSDBNBVRB
35	12.2	67.8	26	5	I26256 Sequence 3
36	12.2	67.8	40	5	AR053640 Sequence
37	12.2	67.8	40	5	AR053647 Sequence
38	12.2	67.8	40	5	I86252 Sequence 6
39	12.2	67.8	40	5	I86255 Sequence 9
40	12.2	67.8	78	5	A37066 Sequence 9
41	12.2	67.8	78	5	A37067 Sequence 10
42	12.2	67.8	83	13	AU025339 Rattus no
43	12.2	67.8	94	13	G38864 TA31 Plasmid
44	12.2	67.8	98	5	I91506 Sequence 40
45	12	66.7	99	10	HSPHA825

## ALIGNMENTS

RESULT 1  
LOCUS I36132 90 bp DNA PAT 13-MAY-1997  
DEFINITION Sequence 16 from patent US 5604131.  
ACCESSION I36132  
VERSION I36132.1 GI:2087356  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 90)  
AUTHORS Wadsworth,S., Snyder,B., Reddy,V.B. and Wei,C.  
TITLE CDNA-genomic DNA hybrid sequence encoding APT70 containing a  
genomic DNA insert of the KI and OX-2 regions  
JOURNAL Patent: US 5604131-A 16-18-FEB-1997;  
FEATURES  
Source 1..90  
Location/Qualifiers  
BASE COUNT 18 a 20 c 12 g 40 t  
ORIGIN

Query Match 80.0%; Score 14.4; DB 5; Length 90;  
Best Local Similarity 93.8%; Pred. No. 2e+03; Mismatches 1; Indels 0; Gaps 0;  
Matches 15; Conservative 0;

OY 3 tttgtgtgtgtgtgt 18  
|||||  
DB 33 TGTGTGTGTGTGT 48

RESULT 2  
LOCUS I27773 36 bp DNA PAT 06-FEB-1997  
DEFINITION Sequence 5 from patent US 5567604.  
ACCESSION I27773  
VERSION I27773.1 GI:1818549  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 36)  
AUTHORS Rando,R.F., Fennewald,S., Zendegeul,J.G. and OJwang,J.O.  
TITLE Anti-viral guanosine-rich oligonucleotides  
JOURNAL Patent: US 5567604-A 5-22-OCT-1996;  
FEATURES  
Source 1..36  
Location/Qualifiers  
BASE COUNT 0 a 0 c 24 g 12 t  
ORIGIN

Query Match 76.7%; Score 13.8; DB 5; Length 36;  
Best Local Similarity 88.2%; Pred. No. 3.8e+03; Mismatches 2; Indels 0; Gaps 0;  
Matches 15; Conservative 0;

OY 1 ggtgtgtgtgtgtgt 17  
|||||  
DB 9 GGTGTGTGTGTGT 25

RESULT 3  
LOCUS I27775 36 bp DNA PAT 06-FEB-1997  
DEFINITION Sequence 7 from patent US 5567604.  
ACCESSION I27775  
VERSION I27775.1 GI:1818551  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 36)

AUTHORS Rando,R.F., Fennewald,S., Zendegeul,J.G. and OJwang,J.O.  
TITLE Anti-viral guanosine-rich oligonucleotides  
JOURNAL Patent: US 5567604-A 7-22-OCT-1996;  
FEATURES  
Source 1..36  
Location/Qualifiers  
BASE COUNT 0 a 0 c 24 g 12 t  
ORIGIN

Query Match 76.7%; Score 13.8; DB 5; Length 36;  
Best Local Similarity 88.2%; Pred. No. 3.8e+03; Mismatches 2; Indels 0; Gaps 0;  
Matches 15; Conservative 0;

OY 1 ggtgtgtgtgtgtgt 17  
|||||  
DB 9 GGTGTGTGTGTGT 25

RESULT 4  
LOCUS I76185 20 bp DNA PAT 03-APR-1998  
DEFINITION Sequence 4 from patent US 5691145.  
ACCESSION I76185  
VERSION I76185.1 GI:3012339  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 20)  
AUTHORS Plitner,J.Bruce, Vonk,G.P. and Nadeau,J.G.  
TITLE Detection of nucleic acids using G-quartets  
JOURNAL Patent: US 5691145-A 4-25-NOV-1997;  
FEATURES  
Source 1..20  
Location/Qualifiers  
BASE COUNT 0 a 0 c 8 g 12 t  
ORIGIN

Query Match 73.3%; Score 13.2; DB 5; Length 20;  
Best Local Similarity 83.3%; Pred. No. 7.6e+03; Mismatches 3; Indels 0; Gaps 0;  
Matches 15; Conservative 0;

OY 1 ggtgtgtgtgtgtgt 18  
|||||  
DB 1 GGTGTGTGTGTGT 18

RESULT 5  
LOCUS I27784 24 bp DNA PAT 06-FEB-1997  
DEFINITION Sequence 16 from patent US 5567604.  
ACCESSION I27784  
VERSION I27784.1 GI:1818560  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 24)  
AUTHORS Rando,R.F., Fennewald,S., Zendegeul,J.G. and OJwang,J.O.  
TITLE Anti-viral guanosine-rich oligonucleotides  
JOURNAL Patent: US 5567604-A 16-22-OCT-1996;  
FEATURES  
Source 1..24  
Location/Qualifiers  
BASE COUNT 0 a 0 c 14 g 10 t  
ORIGIN

Query Match 73.3%; Score 13.2; DB 5; Length 24;  
Best Local Similarity 83.3%; Pred. No. 7.7e+03; Mismatches 3; Indels 0; Gaps 0;  
Matches 15; Conservative 0;

QY 1 ggtgttggtgctt 18  
Db 2 GGTGTGTGTGTGTT 19

## RESULT 6

LOCUS A51237 41 bp DNA PAT 10-MAR-1997

DEFINITION Sequence 8 from Patent WO9614418.

ACCESSION A51237

VERSION A51237.1 GI:2304007

KEYWORDS

SOURCE .

ORGANISM unidentified.

REFERENCE 1 (bases 1 to 41)

AUTHORS Blaz, H., Nguyen, N.T., Stahl, S., Uhlen, M. and Nygren, P. A.

TITLE RESPIRATORY SYNCTIAL VIRUS PROTEIN G EXPRESSED ON BACTERIAL

MEMBRANE

PATENT: WO 9614418-A 8 17-MAY-1996;

PF MEDICAMENT (FR)

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Other publication FR 2726577 960510.

Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ggtgttggtgctt 18  
Db 15 GGTGTGTGTGTGTT 32

## RESULT 8

LOCUS A77180 45 bp DNA PAT 19-OCT-1999

DEFINITION Sequence 8 from Patent EP0614989.

ACCESSION A77180

VERSION A77180.1 GI:6088887

KEYWORDS

SOURCE .

ORGANISM unidentified.

REFERENCE 1 (bases 1 to 45)

AUTHORS Pluckthun, A.D. and Krebber, C.

TITLE A METHOD FOR IN VIVO SELECTION OF LIGAND-BINDING PROTEINS

PATENT: EP 0614989-A 14-SEP-1994;

MORPHOSYS PROTEINOPTIMIERUNG (DE)

Location/Qualifiers

1. 45

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

BASE COUNT 9 a 7 c 13 g 16 t

ORIGIN

Query Match 73.3%; Score 13.2; DB 5; Length 45;  
Best Local Similarity 83.3%; Pred. No. 8e+03;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggtgttggtgctt 18  
Db 15 GGTGTGTGTGTGTT 32

## RESULT 9

LOCUS I20213 45 bp DNA PAT 07-OCT-1996

DEFINITION Sequence 6 from patent US 5514548.

ACCESSION I20213

VERSION I20213.1 GI:1600568

KEYWORDS

SOURCE .

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 45)

AUTHORS Krebber, K., Moroney, S., Pluckthun, A. and Schneider, C.

TITLE Method for in vivo selection of ligand-binding proteins

JOURNAL Patent: US 5514548-A 6 07-MAY-1996;

Location/Qualifiers

1. 45

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

BASE COUNT 9 a 7 c 13 g 16 t

ORIGIN

Query Match 73.3%; Score 13.2; DB 5; Length 45;  
Best Local Similarity 83.3%; Pred. No. 8e+03;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggtgttggtgctt 18  
Db 15 GGTGTGTGTGTGTT 32

## RESULT 10

LOCUS I20215 45 bp DNA PAT 07-OCT-1996

DEFINITION Sequence 8 from patent US 5514548.

ACCESSION I20215

VERSION I20215.1 GI:1600568

KEYWORDS

SOURCE .

ORGANISM Unknown.

BASE COUNT 9 a 7 c 13 g 16 t

ORIGIN

Query Match 73.3%; Score 13.2; DB 5; Length 45;  
Best Local Similarity 83.3%; Pred. No. 8e+03;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggtgttggtgctt 18  
Db 15 GGTGTGTGTGTGTT 32

ACCESSION I20215  
VERSION I20215.1 GI:1600570  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE Unclassified.  
AUTHORS 1 (bases 1 to 45)  
TITLE Kriebber, K., Moroney, S., Pluckhahn, A. and Schneider, C.  
JOURNAL Method for in vivo selection of ligand-binding proteins  
FEATURES Patent: US 5514548-A 8 07-MAY-1996;  
SOURCE Location/Qualifiers  
1..45  
/organism="unknown"  
BASE COUNT 9 a 7 c 13 g 16 t  
ORIGIN

Query Match 73.3%; Score 13.2; DB 5; Length 45;  
Best Local Similarity 83.3%; Pred. No. 8e+03; 3; Indels 0; Gaps 0;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggtgtgtgtgtgctt 18  
||||| ||||| ||| |||  
Db 15 GGTGATGCTGATGATT 32

RESULT 11  
LOCUS A12323 51 bp DNA PAT 06-DEC-1993  
DEFINITION oligonucleotide.  
ACCESSION A12323  
VERSION A12323.1 GI:491330  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 51)  
AUTHORS  
TITLE HYBRID PROTEINS OR POLYPEPTIDES  
JOURNAL Patent: WO 8802757-A 24 21-APR-1988;  
FEATURES Location/Qualifiers  
1..51  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 15 a 27 c 7 g 2 t  
ORIGIN

Query Match 73.3%; Score 13.2; DB 5; Length 51;  
Best Local Similarity 83.3%; Pred. No. 8.1e+03;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggtgtgtgtgtgctt 18  
||||| ||||| ||||| |||  
Db 39 GGCGTTGGGGTTGGCGTT 22

RESULT 12  
LOCUS A12324 51 bp DNA PAT 06-DEC-1993  
DEFINITION oligonucleotide.  
ACCESSION A12324  
VERSION A12324.1 GI:489519  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 51)  
AUTHORS  
TITLE HYBRID PROTEINS OR POLYPEPTIDES  
JOURNAL Patent: WO 8802757-A 25 21-APR-1988;  
FEATURES Location/Qualifiers  
1..51  
source

/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 2 a 7 c 27 g 15 t  
ORIGIN

Query Match 73.3%; Score 13.2; DB 5; Length 51;  
Best Local Similarity 83.3%; Pred. No. 8.1e+03;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggtgtgtgtgtgctt 18  
||||| ||||| ||||| |||  
Db 13 GGCGTTGGGGTTGGCGTT 30

RESULT 13  
LOCUS A12596 51 bp DNA PAT 05-JAN-1994  
DEFINITION oligonucleotide.  
ACCESSION A12596  
VERSION A12596.1 GI:491421  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 51)  
AUTHORS  
TITLE RECOMBINANT VIRUS  
JOURNAL Patent: WO 8701386-A 12 12-MAR-1987;  
FEATURES Location/Qualifiers  
1..51  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 15 a 27 c 7 g 2 t  
ORIGIN

Query Match 73.3%; Score 13.2; DB 5; Length 51;  
Best Local Similarity 83.3%; Pred. No. 8.1e+03;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggtgtgtgtgtgctt 18  
||||| ||||| ||||| |||  
Db 39 GGCGTTGGGGTTGGCGTT 22

RESULT 14  
LOCUS A12597 51 bp DNA PAT 05-JAN-1994  
DEFINITION oligonucleotide.  
ACCESSION A12597  
VERSION A12597.1 GI:489543  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 51)  
AUTHORS  
TITLE RECOMBINANT VIRUS  
JOURNAL Patent: WO 8701386-A 13 12-MAR-1987;  
FEATURES Location/Qualifiers  
1..51  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 2 a 7 c 27 g 15 t  
ORIGIN

Query Match 73.3%; Score 13.2; DB 5; Length 51;  
Best Local Similarity 83.3%; Pred. No. 8.1e+03;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggtgtgtgtgtgctt 18





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:24:08 ; Search time 322.35 Seconds  
(without alignments)

13.971 Million cell updates/sec

Title: US-09-164-714-12

Perfect score: 18  
Sequence: 1 ggcagcaaacaccagctag 18

Scoring table: IDENTITY-NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 431286

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N.Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	13.8	76.7	25	1	063272	Upstream activatio
C 2	13.8	76.7	25	1	081991	Upstream activatio
C 3	13.8	76.7	31	1	063273	Upstream activatio
C 4	13.8	76.7	31	1	063274	Upstream activatio
C 5	13.4	74.4	30	1	054365	Primer/probe D1 sp
C 6	12.8	71.1	22	1	X24317	Rat bone mineralis
C 7	12.4	68.9	30	1	054366	Primer/probe D2 sp
C 8	12.4	68.9	56	1	T25702	Human gene signalu
C 9	12.2	67.8	30	1	V16996	Telomerase competi
C 10	12.2	67.8	48	1	012661	SUC2 signal peptid
C 11	12.2	67.8	55	1	034057	Downstream sequenc
C 12	12.2	67.8	57	1	004915	Wild type signal s
C 13	12.2	67.8	57	1	004916	Modified signal se
C 14	12.2	67.8	57	1	004917	Modified signal se
C 15	12.2	67.8	57	1	004918	Modified signal se
C 16	12.2	67.8	57	1	N81440	Sequence of yeast
C 17	12.2	67.8	63	1	004920	Modified signal se
C 18	12.2	67.8	64	1	043329	Sequence of PCR pr
C 19	12.2	67.8	66	1	004921	Modified signal se
C 20	12.2	67.8	69	1	042522	Sequence encoding
C 21	12.2	67.8	75	1	N71270	Sequence encoding
C 22	12.2	67.8	75	1	011885	Invertase signal -
C 23	12.2	67.8	75	1	011887	Invertase signal -
C 24	12.2	67.8	75	1	N40005	Yeast preinvertase
C 25	12.2	67.8	75	1	N40006	Invertase signal s
C 26	12.2	67.8	75	1	N40008	Invertase signal s
C 27	12.2	67.8	75	1	083672	Partial IGF-I fusi
C 28	12.2	67.8	84	1	083674	Oligo for fusing t
C 29	12.2	67.8	84	1	083673	Partial IGF-I fusi
C 30	12.2	67.8	92	1	056743	Sequence of PCR pr
C 31	12.2	67.8	94	1	T61041	Antisense PCR prim
C 32	11.8	65.6	23	1	T17851	Canola napin promc
C 33	11.8	65.6	29	1	V91720	Human C-Raf hamme
C 34	11.8	65.6	29	1	V91593	Human C-Raf hamme

C 35	11.8	65.6	30	1	054367	Primer/probe D3 sp
C 36	11.8	65.6	30	1	054368	Primer/probe D4 sp
C 37	11.8	65.6	30	1	054369	Primer/probe D5 sp
C 38	11.8	65.6	40	1	030844	Type III procollag
C 39	11.8	65.6	81	1	073051	Conserved sequence
C 40	11.6	64.4	21	1	X15036	Antisense PCR prim
C 41	11.6	64.4	21	1	X15012	Probe used to isol
C 42	11.6	64.4	37	1	049410	PCR primer for amp
C 43	11.6	64.4	42	1	X16930	Primer 028 for con
C 44	11.4	63.3	17	1	V97277	Human EGF-R target
C 45	11.4	63.3	17	1	V97278	Human EGF-R target

## ALIGNMENTS

RESULT 1  
063272/c  
ID 063272 standard; cDNA: 25 BP.  
AC 063272;  
DT 14-DEC-1994 (first entry)  
DE Upstream activation sequence of AOX2 promoter.  
KW AOX2; Alcohol oxygenase; Promoter region; enhanced activity; UAS;  
KW upstream activation sequence; heterologous protein expression; ss.  
OS Pichia pastoris.  
PN EP-595334-A.  
PD 04-MAY-1994.  
PE 28-OCT-1993; 117527.  
PR 30-OCT-1992; JP-293315.  
PR 06-AUG-1993; JP-215306.  
PA (GREG ) GREEN CROSS CORP.  
PI Chuganji M, Hiramatsu R, Miura M, Ohi H, Ohmura T;  
PI Uno S, Hiramatsu F;  
DR WPI: 94-145941/18.  
PT New mutant AOX promoter - having a deletion, replacement or  
insertion of nucleotide to increase transcription activity in  
PT prodn. of proteins  
PS Example 11; Page 13; 48pp; English.  
CC 063272 is a putative UAS of AOX2 promoter.  
CC The promoter is mutated such that it has an enhanced activity. The  
CC mutant AOX2 promoter is useful in an expression vector allowing  
CC heterologous protein expression.  
SQ Sequence 25 BP; 7 A; 3 C; 5 G; 10 T;  
  
Query Match 76.7%; Score 13.8; DB 1; Length 25;  
Best local similarity 88.2%; Pred. No. 1.1e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 ggcagcaaacaccagctag 17  
Db 19 GCGACAAATAACCTTA 3  
  
RESULT 2  
081991/c  
ID 081991 standard; DNA: 25 BP.  
AC 081991;  
DT 28-SEP-1995 (first entry)  
DE Upstream activation sequence for mutant alcohol oxidase promoter.  
KW Alcohol oxidase promoter; upstream activation sequence; ss.  
OS Synthetic.  
PN EP-639643-A.  
PD 22-FEB-1995.  
PR 26-JUL-1994; 111628.  
PR 27-JUL-1993; JP-185003.  
PA (GREG ) GREEN CROSS CORP.  
PI Hiramatsu R, Miura M, Ohi H, Ohmura T;  
PI WPI: 95-083455/12.  
PT New mutant alcohol oxidase promoter - for heterologous protein  
expressions esp. in yeast, has at least one added sequence  
PT enhancing transcriptional activity  
Claim 1; Page 18; 40pp; English.

CC Q81991 is an upstream activation sequence which is added in the  
CC reverse direction or in both the forward and reverse directions at  
CC the 5' end of a partial DNA fragment of a wild-type alcohol oxidase  
CC (AOX2) promoter. The sequence of the wild-type AOX2 promoter is  
CC given in Q81992. The partial fragment of AOX2 is indicated in FT  
CC Q81992. The partial fragment also pref. has certain mutations, also  
CC indicated in FT. Incorporation of the upstream activation sequence  
CC provides more efficient expression of heterologous downstream genes.  
SQ Sequence 25 BP; 7 A; 3 C; 5 G; 10 T;

Query Match 76.7%; Score 13.8; DB 1; Length 25;  
Best Local Similarity 88.2%; Pred. No. 1.1e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ggcacaaacagccta 17  
|||||  
DB 19 GCGACAAATAGCCTA 3

RESULT 3  
063273/c  
ID 063273; standard; cDNA; 31 BP.  
AC 063273;  
DT 15-DEC-1994 (first entry)  
DE Upstream activation sequence of AOX2 promoter.  
KM AOX2; Alcohol oxygenase; Promoter region; enhanced activity; UAS;  
KM upstream activation sequence; heterologous protein expression; ss.  
OS Pichia pastoris.  
PH Key Location/Qualifiers  
FT enhancer 6..31  
FT /tag= a  
FT /note= "putative upstream activation sequence"

EP-595334-A.  
PD 04-MAY-1994.  
PR 28-OCT-1993; 117527.  
PR 30-OCT-1992; JP-293315.  
PR 06-AUG-1993; JP-215306.  
PA (GRNC) GREEN CROSS CORP.  
PI Chuganji M, Hiramatsu R, Miura M, Ohl H, Ohmura T;  
PI Uno S, Hiramatsu F;  
DR WPI: 94-145941/18.  
PR New mutant AOX promoter - having a deletion, replacement or  
PR insertion of nucleotide to increase transcription activity in  
PR prodn. of proteins  
PS Example 11; Page 13; 48pp; English.  
CC 063272 is a putative UAS of AOX2 promoter from Pichia pastoris. The  
CC 25bp sequence was chemically synthesised and EcoRI sites added.  
CC The promoter is mutated such that it has an enhanced activity. The  
CC mutant AOX2 promoter is useful in an expression vector allowing  
CC heterologous protein expression.  
SQ Sequence 31 BP; 9 A; 4 C; 6 G; 12 T;

Query Match 76.7%; Score 13.8; DB 1; Length 31;  
Best Local Similarity 88.2%; Pred. No. 1.1e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ggcacaaacagccta 17  
|||||  
DB 24 GCGACAAATAGCCTA 8

RESULT 4  
063274  
ID 063274; standard; cDNA; 31 BP.  
AC 063274;  
DT 15-DEC-1994 (first entry)

DE Upstream activation sequence of AOX2 promoter.  
KM AOX2; Alcohol oxygenase; Promoter region; enhanced activity; UAS;  
KM upstream activation sequence; heterologous protein expression; ss.  
OS Pichia pastoris.  
PN EP-595334-A.

PD 04-MAY-1994.  
PR 28-OCT-1993; 117527.  
PR 30-OCT-1992; JP-293315.  
PR 06-AUG-1993; JP-215306.  
PA (GRNC) GREEN CROSS CORP.  
PI Chuganji M, Hiramatsu R, Miura M, Ohl H, Ohmura T;  
PI Uno S, Hiramatsu F;  
DR WPI: 94-145941/18.

PT New mutant AOX promoter - having a deletion, replacement or  
PT insertion of nucleotide to increase transcription activity in  
PT prodn. of proteins  
PS Example 11; Page 13; 48pp; English.  
CC 063273 is a putative UAS of AOX2 promoter from Pichia pastoris. The  
CC 25bp sequence (reverse strand) was chemically synthesised and EcoRI  
CC sites added. The promoter is mutated such that it has an enhanced  
CC activity. The mutant AOX2 promoter is useful in an expression vector  
CC allowing heterologous protein expression.  
SQ Sequence 31 BP; 12 A; 6 C; 4 G; 9 T;

Query Match 76.7%; Score 13.8; DB 1; Length 31;  
Best Local Similarity 88.2%; Pred. No. 1.1e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ggcacaaacagccta 17  
|||||  
DB 12 GCGACAAATAGCCTA 28

RESULT 5  
054365/c  
ID 054365; standard; DNA; 30 BP.  
AC 054365;  
DT 24-JUN-1994 (first entry)  
DE Primer/probe D1 specific for Erwinia chrysanthemi.  
KM Erwinia chrysanthemi strain EC16; pectinolytic bacteria; pelE gene;  
KM polymerase chain reaction amplification; detection; pectate lyase;  
KM plant pathogen; ss.  
OS Erwinia chrysanthemi (strain EC16).  
PN WO9325708-A.  
PD 23-DEC-1993.  
PR 04-JUN-1993; F00540.  
PR 05-JUN-1992; FR-006888.  
PA (INRA-) INRA INST NAT AGRONOMIQUE PARIS-GRIGNON.  
PA (INRG) INRA INST NAT RECH AGRONOMIQUE.  
PI Darresse A, Kotoujansky A, Bertheau Y;  
DR WPI: 94-007565/01.  
PR Identification and detection of specific Erwinia species,  
PR sub-species and pathovars - by amplifying pectate lyase DNA then  
PR hybridisation testing  
PS Claim 5; Page 33; 47pp; French.  
CC Probe/primer D1 is one of 5 preferred oligonucleotides derived from  
CC the sequence complementary to nucleotides 672-701 of the pectate  
CC lyase gene pelE from E. chrysanthemi strain 16. The oligonucleotide  
CC is used with probe/primer C (054362-054364) to specifically amplify  
CC a region of the pel gene. Early diagnosis of plant diseases caused  
CC by E. chrysanthemi is possible using the primers.  
CC See 054371 for the full-length pelE gene.  
SQ Sequence 30 BP; 3 A; 6 C; 11 G; 10 T;

Query Match 74.4%; Score 13.4; DB 1; Length 30;  
Best Local Similarity 93.3%; Pred. No. 1.8e+02;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 gcacaaacagcct 16  
|||||  
DB 27 GCGACAAACATCCT 13

RESULT 6  
X24317/c  
ID X24317; standard; DNA; 22 BP.

AC X24317;  
 DT 07-JUN-1999 (first entry)  
 DE Rat bone mineralisation protein RIMP reverse PCR primer.  
 KW RIMP; LMP; LIM mineralisation protein; bone mineralisation; rat;  
 KW cytokine; spine fusion; fracture repair; bone grafting;  
 KW osteoporosis; gene therapy; PCR; primer; ss.  
 OS Synthetic.  
 OS Rattus norvegicus.  
 PN W0906563-A1.  
 PD 11-FEB-1999.  
 PF 29-JUL-1998; U15814.  
 PR 02-APR-1998; US-080407.  
 PR 30-JUL-1997; US-054219.  
 PA (UYEM-) UNTV EMORY.  
 PI Boden S, Hair G;  
 DR WPI: 99-153793/13.  
 PT LIM mineralisation proteins - used to induce bone formation from  
 PT osteogenic precursors, e.g. for treating osteoporosis or repairing  
 PT fractures  
 PS Example 16: Page 53; 67pp; English.  
 CC This reverse primer, and a forward primer (see X24316), are based  
 CC on rat LIM mineralisation protein (RIMP) cDNA (see X24307).  
 CC The primers were used to amplify a unique 223 bp sequence from  
 CC rat RIMP cDNA. A similar PCR product (see X24319) was isolated  
 CC from human MG63 osteosarcoma cell cDNA with the same PCR primers.  
 CC The human amplified sequence showed 95% homology to the amplified  
 CC rat sequence. The reverse primer was also used with a different  
 CC specific primer (see X24318) in a PCR amplification of MG63 cDNA  
 CC to obtain another partial sequence (see X24320). A full-length  
 CC human LMP cDNA (see X24322) was subsequently obtained. The  
 CC encoded protein (see M97844) enhances bone mineralisation in vitro  
 CC and in vivo. Clinical applications include enhancement of bone  
 CC repair in fractures, bone defects, bone grafting, and normal  
 CC homeostasis in patients presenting with osteoporosis. Methods of  
 CC inducing bone formation using transfected osteogenic precursor  
 CC cells are provided.  
 SQ Sequence 22 BP; 1 A; 4 C; 8 G; 9 T;

Query Match 71.1%; Score 12.8; DB 1; Length 22;  
 Best Local Similarity 87.5%; Pred. No. 3.5e+02;  
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 gacaaacagcctag 18  
 |||||||||  
 DB 19 GACAAACGACGACAG 4

RESULT 7  
 O54366/c  
 ID 054366 standard; DNA; 30 BP.  
 AC 054366;  
 DT 24-JUN-1994 (first entry)  
 DE Primer/probe D2 specific for Erwinia chrysanthemi.  
 KW Erwinia chrysanthemi strain EC16; pectinolytic bacterium; pelf gene;  
 KW polymerase chain reaction amplification; detection; pectate lyase;  
 KW plant pathogen; ss.  
 OS Erwinia chrysanthemi (strain EC16).  
 OS W09325708-A.  
 PN W09325708-A.  
 PD 23-DEC-1993.  
 PF 04-JUN-1993; F00540.  
 PR 05-JUN-1992; FR-006888.  
 PA (INRA-) INRA INST NAT AGRONOMIQUE PARIS-GRIGNON.  
 PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
 PI Darrasse A, Kotoujansky A, Bertheau Y;  
 DR WPI: 94-007565/01.  
 PT Identification and detection of specific Erwinia species,  
 PT sub-species and pathogens - by amplifying pectate lyase DNA then  
 PT hybridisation testing  
 PT Claim 5; Page 33; 47pp; French.  
 CC Probe/primer D2 is one of 5 preferred oligonucleotides derived from  
 CC the sequence complementary to nucleotides 672-701 of the pectate  
 CC lyase gene pelf from E.chrysanthemi strain 16. The oligonucleotide

CC is used with probe/primer C (O54362-O54364) to specifically amplify  
 CC a region of the pelf gene. Early diagnosis of plant diseases caused  
 CC by E.chrysanthemi is possible using the primers.  
 CC See 054371 for the full-length pelf gene.  
 SQ Sequence 30 BP; 5 A; 6 C; 9 G; 10 T;

Query Match 68.9%; Score 12.4; DB 1; Length 30;  
 Best Local Similarity 92.9%; Pred. No. 5.8e+02;  
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 gacaaacagcct 16  
 |||||||||  
 DB 26 GACAAACGACATCCT 13

RESULT 8  
 T25702/c  
 ID T25702 standard; cDNA to mRNA; 56 BP.  
 AC T25702;  
 DT 10-OCT-1996 (first entry)  
 DE Human gene signature HUMGS07903.  
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
 KW human; cloning; mapping; non-biased library; diagnosis; detection;  
 KW cell typing; abnormal cell function; ss.  
 OS Homo sapiens.  
 OS W09514772-A1.  
 PN W09514772-A1.  
 PD 01-JUN-1995.  
 PF 11-NOV-1994; J01916.  
 PR 12-NOV-1993; JP-355504.  
 PA (MATS/) MATSUBARA K.  
 PA (OKUB/) OKUBO K.  
 PI Matsubara K, Okubo K;  
 DR WPI: 95-206931/27.  
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
 PT for diagnosis of abnormal cell function, by preparing cDNA that  
 PT reflects relative abundance of corresp. mRNA in specific human  
 PT tissues  
 PS Claim 1; Page 1910; 2245pp; Japanese.  
 CC A single-stranded DNA (or its complementary strand or the corresp.  
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
 CC given in T19001-T26837 and which is able to hybridise to part of  
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
 CC sequences were obtained from 3'-directed cDNA libraries prepared  
 CC from various human tissues; synthesis of cDNA was initiated from the  
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
 CC untranslated sequence is unique to a particular mRNA species, almost  
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
 CC is constructed so as to reflect accurately the relative abundance of  
 CC different mRNAs in the particular tissue from which it was derived.  
 CC The appearance frequency of a given GS in a cDNA library can be  
 CC determined (esp. using primers and probes derived from the GS  
 CC sequences) as a means of diagnosing abnormal cell function or for  
 CC recognising different cell types.  
 SQ Sequence 56 BP; 13 A; 15 C; 16 G; 12 T;

Query Match 68.9%; Score 12.4; DB 1; Length 56;  
 Best Local Similarity 92.9%; Pred. No. 6.1e+02;  
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 caaaacagcctag 18  
 |||||||||  
 DB 32 CAAAAACGACCCAG 19

RESULT 9  
 V16996/c  
 ID V16996 standard; DNA; 30 BP.  
 AC V16996;  
 DT 13-AUG-1998 (first entry)  
 DE Telomerase competitor DNA oligonucleotide.  
 KW Human; telomerase reverse transcriptase; hTERT; TRT; diagnosis;

KW prognosis; cell proliferation; cancer; ageing; ribonucleoprotein; ss.  
 OS Synthetic.  
 PN GB317891-A.  
 PD 08-APR-1998.  
 PE 01-OCT-1997; 020890.  
 PR 14-AUG-1997; US-915503.  
 PR 01-OCT-1996; US-724643.  
 PR 18-APR-1997; US-844419.  
 PR 25-APR-1997; US-846017.  
 PR 06-MAY-1997; US-851843.  
 PR 09-MAY-1997; US-854050.  
 PR 14-AUG-1997; US-911312.  
 PR 14-AUG-1997; US-912951.  
 PA (GERO-) GERON CORP.  
 PA (UYTE-) UNIV TECHNOLOGY CORP.  
 PI Andrews WH, Czech TR, Chapman KB, Harley C, Lingner J,  
 PI Morin GB, Nakamura T, Harley CB;  
 DR WPI: 98-171633/16.  
 PT Pure and recombinant human Telomerase Reverse Transcriptase and its  
 PT variants - are useful in the diagnosis, prognosis and treatment of  
 PT cell proliferation conditions especially cancer and ageing  
 PS Example 1; Page 190; 387pp; English.  
 CC The present sequence represents a competitor DNA oligonucleotide  
 CC used in an example from the present invention which describes  
 CC human telomerase reverse transcriptase (hTERT). The present  
 CC invention also describes the following methods: (A) determining  
 CC whether a test compound is a modulator of hTERT, by detecting the change  
 CC in hTERT recombinant protein or polynucleotide, on administration of the  
 CC compound; (B) preparation of recombinant telomerase by contacting a  
 CC protein preparation of hTERT with a telomerase RNA component; (C)  
 CC detection of the hTERT RNA or protein in a sample by binding a relevant  
 CC probe to the sample and detecting the complex formed or in the case of  
 CC RNA detection, amplifying the product and correlating the presence of  
 CC complex or amplification product with presence of hTERT in the sample;  
 CC and (D) increasing the proliferation of a vertebrate cell by increasing  
 CC hTERT expression; and (E) the use of an agent that causes an increase in  
 CC cell vertebrate cell proliferation to create a medicament that inhibits  
 CC ageing. A protein preparation of hTERT and the polynucleotide encoding  
 CC hTERT can be used in the manufacture of medicaments for inhibiting the  
 CC effect of ageing or cancer. Inhibitors of telomerase activity can be  
 CC used to treat conditions that are associated with high telomerase  
 CC activity. A protein preparation of hTERT can also be used in the new  
 CC methods.  
 SQ Sequence 30 BP; 6 A; 4 C; 9 G; 11 T;

Query Match 67.8%; Score 12.2; DB 1; Length 30;  
 Best Local Similarity 82.4%; Pred. No. 7.3e+02;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ggcacaaacagccta 17  
 ||||| |||||  
 DB 26 GCCACAAATCCACCTA 10

RESULT 10  
 Q12661/c  
 ID Q12661 standard; DNA; 48 BP.  
 AC Q12661:  
 DT 01-OCT-1991 (first entry)  
 DE SUC2 signal peptide.  
 KW sucrose invertase; yeast; *Pichia pastoris*; mutant MP-36; ds.  
 OS Synthetic.  
 PN EP-438200-A.  
 PD 24-JUL-1991.  
 PR 16-JAN-1991; 200074.  
 PR 16-JAN-1990; CU-000007.  
 PR 13-JUN-1990; CU-000104.  
 PR 12-JUL-1990; CU-000132.  
 PR 15-AUG-1990; CU-000142.  
 PA (INGE-) CENT INGEN GENETICA.  
 PI Herrera-Martinez LS, Yonng-Gonzalez V, Margolles-Clark E,  
 PI Delgado-Boada JM, Morales-Grillo J, Torres-Madrado IC,

PI Silva-Rodriguez A, Palfer-Reyes E;  
 DR WPI: 91-216886/30.  
 PT Expression of heterologous genes in *Pichia pastoris* yeast - by  
 PT using expression vectors and transformed microorganisms  
 PS Example 2; Fig 5; 24pp; English.  
 CC The 5' end overlaps the 3' end of the complementary strand by 4  
 CC bases; the 5' end of the complementary strand overlaps the 3' end  
 CC of the sense strand by the four bases 5'CATG 3'. The cohesive ends  
 CC are HindIII and NcoI sites, respectively. The synthetic sequence is  
 CC cloned into a yeast expression vector to give plasmid pps-7 which  
 CC contains the AOX1 promoter, the SUC2 signal sequence, a fragment of  
 CC the *S.cerevisiae* GAP operator (GAP), the *S.cerevisiae* HIS3 gene  
 CC and the 3' terminal fragment of the AOX gene for integration in the  
 CC yeast. Insertion of an heterologous gene occurs between the SUC2  
 CC signal sequence and GAP using the regenerated NcoI site.  
 CC See also Q12662.  
 SQ Sequence 48 BP; 8 A; 12 C; 10 G; 18 T;

Query Match 67.8%; Score 12.2; DB 1; Length 48;  
 Best Local Similarity 82.4%; Pred. No. 7.7e+02;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ggcacaaacagccta 17  
 ||||| |||||  
 DB 32 GCTGCAAAACGACCA 16

RESULT 11  
 Q34057  
 ID Q34057 standard; DNA; 55 BP.  
 AC Q34057:  
 DT 02-FEB-1993 (first entry)  
 DE Downstream sequence of microsatellite from clone TGM44.  
 KW PCR; selection; primers; OPTIRIM; breeding; cattle; parentage;  
 KW genetic mapping; traits; amplification; ss.  
 OS Bos taurus.  
 PN WO9213102-A.  
 PD 06-AUG-1992.  
 PR 15-JAN-1992; U00340.  
 PR 15-JAN-1991; US-642342.  
 PA (GENM-) GENMARK.  
 PI Georges M, Massey JM;  
 DR WPI: 92-284684/34.  
 PT Polymorphic bovine DNA markers - used in genetic identification,  
 PT gene mapping, and selective breeding  
 PS Table 7; Page 353; 517pp; English.  
 CC The sequence is that downstream of a bovine microsatellite sequence  
 CC obtd. by screening a library of bovine MboI DNA fragments of between  
 CC 250 and 500 bp with an (AC)<sub>15</sub> and a (TC)<sub>15</sub> oligonucleotide probe.  
 CC One out of 50 clones cross-hybridised. Assuming independent  
 CC distribution of microsatellites and MboI sites, the frequency of  
 CC (T6)n >9 microsatellites in the bovine genome is estimated at >100.  
 CC 000. The sequence information for ca. 230 such bovine microsatellites  
 CC is summarised in the specification and indexed herein (see below).  
 CC The sequences upstream and downstream of the microsatellite sequence  
 CC were used to generate the required PCR primers for in vitro  
 CC amplification of the corresp. microsatellite (using the program  
 CC OPTIRIM). The microsatellites may be used to identify individuals,  
 CC for parentage testing, and in the genetic mapping of economic trait  
 CC loci, or genes involved in the determination of economically important  
 CC traits esp. in cattle, to allow selective breeding.  
 CC See also Q33501-34437.  
 SQ Sequence 55 BP; 19 A; 16 C; 11 G; 9 T;

Query Match 67.8%; Score 12.2; DB 1; Length 55;  
 Best Local Similarity 82.4%; Pred. No. 7.8e+02;  
 Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ggcacaaacagccta 17  
 ||||| |||||  
 DB 37 GCGACTAAACACCATTA 53

## RESULT 12

004915/c  
ID 004915 standard; DNA; 57 BP.  
AC 004915;  
DT 23-OCT-1990 (first entry)  
DE Wild type signal sequence of plasmid pING156.  
KW Thaumatin; sweetener; sugar; sucrose; ds.  
OS Thaumatooccus daniellii.  
PN W09005775-A.  
PD 31-MAY-1990.  
PF 6-NOV-1989; 005018.  
PR 8-NOV-1988; US-268702.  
PS 4-SEP-1989; US-407416.  
PT (ITGE-) Int Genetics Eng Inc.  
PI Blair LC, Koduri RK, Lee JH, Weickmann JL;  
DR WPI: 90-193426/25.  
P-PSDB: R05497.  
PT DNA encoding thaumatin 1 analogues with improved taste -  
PT for use as sweeteners and tools for investigating the taste  
response.  
PS Example 5; Table 5; 65pp; English.  
CC Thaumatin is 5000 times sweeter in taste than sucrose, and has  
potential as sugar substitute. The plasmid contains promoter,  
secretion signal and terminator sequences, and may be used in an  
E.coli or S.cerevisiae expression system.  
CC The modified signal sequence increases yield and decreases lag  
time for the system.  
SQ Sequence 57 BP; 9 A; 14 C; 10 G; 24 T;

Query Match 67.8%; Score 12.2; DB 1; Length 57;  
Best Local Similarity 82.4%; Pred. No. 7.8e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ggcacaaaccagccta 17  
||| ||||| ||||| |||  
DB 44 GCTGCAAAACGACCA 28

## RESULT 13

004916/c  
ID 004916 standard; DNA; 57 BP.  
AC 004916;  
DT 23-OCT-1990 (first entry)  
DE Modified signal sequence of plasmid pING447.  
KW Thaumatin; sweetener; sugar; sucrose; ds.  
OS Thaumatooccus daniellii.  
PN W09005775-A.  
PD 31-MAY-1990.  
PF 6-NOV-1989; 005018.  
PR 8-NOV-1988; US-268702.  
PS 4-SEP-1989; US-407416.  
PT (ITGE-) Int Genetics Eng Inc.  
PI Blair LC, Koduri RK, Lee JH, Weickmann JL;  
DR WPI: 90-193426/25.  
P-PSDB: R05498.  
PT DNA encoding thaumatin 1 analogues with improved taste -  
PT for use as sweeteners and tools for investigating the taste  
response.  
PS Example 5; Table 5; 65pp; English.  
CC Thaumatin is 5000 times sweeter in taste than sucrose, and has  
potential as sugar substitute. The plasmid contains promoter,  
secretion signal and terminator sequences, and may be used in an  
E.coli or S.cerevisiae expression system.  
CC The modified signal sequence increases yield and decreases lag  
time for the system.  
SQ Sequence 57 BP; 11 A; 13 C; 11 G; 22 T;

Query Match 67.8%; Score 12.2; DB 1; Length 57;  
Best Local Similarity 82.4%; Pred. No. 7.8e+02;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ggcacaaaccagccta 17  
||| ||||| ||||| |||  
DB 44 GCTGCAAAACGACCA 28

## RESULT 14

004917/c  
ID 004917 standard; DNA; 57 BP.  
AC 004917;  
DT 23-OCT-1990 (first entry)  
DE Modified signal sequence of plasmid pING448.  
KW Thaumatin; sweetener; sugar; sucrose; ds.  
OS Thaumatooccus daniellii.  
PN W09005775-A.  
PD 31-MAY-1990.  
PF 6-NOV-1989; 005018.  
PR 8-NOV-1988; US-268702.  
PS 4-SEP-1989; US-407416.  
PT (ITGE-) Int Genetics Eng Inc.  
PI Blair LC, Koduri RK, Lee JH, Weickmann JL;  
DR WPI: 90-193426/25.  
P-PSDB: R05499.  
PT DNA encoding thaumatin 1 analogues with improved taste -  
PT for use as sweeteners and tools for investigating the taste  
response.  
PS Example 5; Table 5; 65pp; English.  
CC Thaumatin is 5000 times sweeter in taste than sucrose, and has  
potential as sugar substitute. The plasmid contains promoter,  
secretion signal and terminator sequences, and may be used in an  
E.coli or S.cerevisiae expression system.  
CC The modified signal sequence increases yield and decreases lag  
time for the system.  
SQ Sequence 57 BP; 11 A; 13 C; 11 G; 22 T;

Query Match 67.8%; Score 12.2; DB 1; Length 57;  
Best Local Similarity 82.4%; Pred. No. 7.8e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ggcacaaaccagccta 17  
||| ||||| ||||| |||  
DB 44 GCTGCAAAACGACCA 28

## RESULT 15

004918/c  
ID 004918 standard; DNA; 57 BP.  
AC 004918;  
DT 23-OCT-1990 (first entry)  
DE Modified signal sequence of plasmid pING449.  
KW Thaumatin; sweetener; sugar; sucrose; ds.  
OS Thaumatooccus daniellii.  
PN W09005775-A.  
PD 31-MAY-1990.  
PF 6-NOV-1989; 005018.  
PR 8-NOV-1988; US-268702.  
PS 4-SEP-1989; US-407416.  
PT (ITGE-) Int Genetics Eng Inc.  
PI Blair LC, Koduri RK, Lee JH, Weickmann JL;  
DR WPI: 90-193426/25.  
P-PSDB: R05500.  
PT DNA encoding thaumatin 1 analogues with improved taste -  
PT for use as sweeteners and tools for investigating the taste  
response.  
PS Example 5; Table 5; 65pp; English.  
CC Thaumatin is 5000 times sweeter in taste than sucrose, and has  
potential as sugar substitute. The plasmid contains promoter,  
secretion signal and terminator sequences, and may be used in an  
E.coli or S.cerevisiae expression system.  
CC The modified signal sequence increases yield and decreases lag  
time for the system.

Sequence 57 BP; 10 A; 12 C; 11 G; 24 T;

Query Match 67.8%; Score 12.2; DB 1; Length 57;

Best Local Similarity 82.4%; Pred. No. 7.8e+02;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 gcgacaaaccagccta 17

Db 44 GCTGCAAAACGAGCCAA 28

Search completed: June 4, 2000, 16:24:10  
Job time: 28906 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:24:15 ; Search time 322.35 Seconds  
(without alignments)  
34,927 Million cell updates/sec

Title: US-09-164-714-16

Perfect score: 45  
Sequence: 1 atcaagctagatgagtgtga.....atgaaagccaatgagcgc 45

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 431286

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28.8	64.0	48	1	T73188 Mycoplasma arthrit
2	25.6	56.9	45	1	069019 E. coli tryptophan
3	25.6	56.9	45	1	069017 E. coli tryptophan
4	25.6	56.9	42	1	069013 Bacteriophage M13
5	25.6	55.1	54	1	069028 Bacteriophage M13
6	24.8	55.1	75	1	V84716 Monoclonal antibod
7	24.6	54.7	69	1	V93934 Oligonucleotide us
8	24.6	54.7	69	1	V93936 Oligonucleotide us
9	24.4	54.2	55	1	X07146 Staphylococcus aur
10	24.4	54.2	55	1	X07148 Staphylococcus aur
11	24.2	53.8	69	1	V29677 Bacillus termamyl
12	24.2	53.3	48	1	X32832 Seq ID No:16 of JP
13	23.2	51.6	62	1	V86012 Mouse LRP-3 cDNA P
14	23.2	51.1	51	1	079619 Encodes Factor Xa
15	22.6	50.2	58	1	V81820 Granulocyte Ehrli
16	22.4	49.8	39	1	V69813 Human fibronectin
17	22.4	49.8	53	1	T64581 Primer for thymus
18	22.4	49.8	53	1	V31475 PCR primer 3'-AP(H
19	22.4	49.8	53	1	V38477 Human CC chemokine
20	22.4	49.8	63	1	T13818 T-cell receptor al
21	22.2	49.3	62	1	V40242 Mouse VEGF-B186 ps
22	22.2	48.9	38	1	V07490 Oligonucleotide co
23	22.2	48.9	38	1	V07491 Oligonucleotide co
24	21.8	48.4	37	1	069020 E. coli tryptophan
25	21.6	48.0	60	1	T29761 B10 scfcr hexahist
26	21.2	47.1	51	1	V44009 Human Mab #117/10C
27	21.2	47.1	51	1	T66453 Primer for human p
28	21.2	46.7	47	1	Q87429 Human GSK catalyti
29	21.1	46.7	50	1	X24790 Interleukin-18 bin
30	20.4	45.3	89	1	V84718 Monoclonal antibod
31	20.4	45.3	42	1	T37396 Plasmid pOMP2-5 f
32	20.4	45.3	42	1	T37397 Plasmid pOMP2-5 f
33	20.4	45.3	43	1	V56674 Human ELL2 cDNA PC
34	20.4	45.3	50	1	V50374 PCR primer for KDR

C 35	20.4	45.3	51	1	V47567 PCR primer for let
C 36	20.4	45.3	52	1	T49499 His tag PCR primer
C 37	20.4	45.3	63	1	Q70464 pEX-6Hpro proteas
C 38	20.4	45.3	69	1	V04222 Modified human car
C 39	20.4	45.3	72	1	X25976 Human IL-5 gene pr
C 40	20.2	44.9	48	1	T14088 Human parvovirus B
C 41	20.2	44.4	56	1	T00754 Primer for extrace
C 42	20	44.4	99	1	V39623 pFcl DNA oligonuc
C 43	19.8	44.0	27	1	X00837 Insert sequence H1
C 44	19.8	44.0	58	1	X23302 Human SPA-1 PCR pr
C 45	19.8	44.0	66	1	X23303 Human SPA-1 PCR pr

## ALIGNMENTS

RESULT 1					
ID T73188	standard; DNA; 48 BP.				
AC T73188;					
DT 27-MAR-1998	(first entry)				
DE Mycoplasma arthritidis T cell mitogen gene PCR primer.					
KW T cell mitogen; mam gene; superantigen; antigen;					
KW rheumatoid arthritis; diagnosis; PCR; primer; ss.					
OS Synthetic.					
OS Mycoplasma arthritidis PC6.					
PN WO9734913-A1.					
PD 25-SEP-1997.					
PF 21-MAR-1997; U05067.					
PR 22-MAR-1996; US-621081.					
PA (UTAH ) UNIV UTAH RES FOUND.					
PI Atkin CL, Cole BC, Knudtson KL, Sawitzke AD;					
DR WPI; 97-480155/44.					
PT Recombinant Mycoplasma arthritidis T cell mitogen protein					
PT exhibiting superantigen activity; useful to diagnose rheumatoid					
PT arthritis					
PS Example 10; Page 26; 66pp; English.					
CC PCR primers (T73187 and T73188) were designed to amplify					
CC a recombinant Mycoplasma arthritidis T cell mitogen (MAM) gene					
CC (see T73178) derivative with an in-frame motif coding for					
CC C-terminal hexahistidine tag. An expression vector for					
CC maltose binding protein-MAM fusion was used as template. The					
CC PCR product was ligated to plasmid pQE60 to generate pKK22.					
CC Recombinant MAM protein having a C-terminal hexahistidine					
CC tag (see W23459) was produced in E. coli transformants. The					
CC recombinant MAM exhibits superantigen activity and can be used in					
CC the diagnosis of rheumatoid arthritis.					
SO Sequence 48 BP; 13 A; 6 C; 13 G; 16 T;					
Query Match	64.0%; Score 28.8; DB 1; Length 48;				
Best Local Similarity	82.5%; Pred. No. 0.037;				
Matches 33; Conservative 0; Mismatches 7; Indels 0; Gaps 0;					
Qy 4 aagcttagtggatggtgagtgatgaaagccaatgagc 43					
Db 3 AAGCTTAGTGATGGATGGATGATCTTCAAAAACAGC 42					
RESULT 2					
ID 069019	standard; DNA; 45 BP.				
AC 069019;					
DT 11-APR-1995	(first entry)				
DE E. coli tryptophan synthase beta-subunit primer #4.					
KW Polymerase chain reaction; primer; amplify; PCR; long; short;					
KW bacteriophage M13; gene III; E. coli; tryptophan synthase;					
KW beta-subunit; fusion sequence; ligand binding protein; antibody; ss.					
OS Synthetic.					
PN EP-614989-A.					
PD 14-SEP-1994.					
PF 16-FEB-1994; 102334.					
PR 17-FEB-1993; EP-102484.					

PA (MORP- MORPHOSIS GES PROTEINOPTIMIERUNG MBH.  
PI Kriebler C, Moroney S, Plueckthun A, Schneider C;  
DR WIR; 94-281214/35.  
PT Selection of genes encoding ligand binding proteins - using  
PT replicable genetic packages infective only in presence of binding  
PT protein having high affinity for target ligand  
PS Example 1; Page 8; 26pp; English.  
CC The sequences given in Q69012-15 are primers which were used to  
CC amplify the long and short versions of the bacteriophage M13 genome  
CC III. The amplified sequence was ligated to the sequence amplified  
CC by the primer sequences given in Q69016-19, which encodes the E.  
CC typhlophan synthase beta-subunit. When the fusion sequence is  
CC expressed it may be used to isolate ligand binding proteins, eg.  
CC antibodies which have a high affinity for a specific ligand, esp.  
CC from a large collection of valient molecules.  
SO Sequence 45 BP; 9 A; 7C; 13 G; 16 T;

Query Match	56.98;	Score 25.6;	DB 1;	Length 45;
Best Local Similarity	87.58;	Pred. No. 0.51;		
Matches 28;	Conservative	0;	Mismatches 4;	Indels 0;
				Gaps 0

QY	4	aagcttagtga	tgtgatgta	aaagcc	35
Db	3	AAGCTTAGTGA	TGCGTGATG	GATTTC	34

RESULT	3
Q69017	
ID	Q69017 standard; DNA; 45 BP

DT 11-APR-1995 (first entry)  
DE E. coli tryptophan synthase beta-subunit primer #2.  
KW polymerase chain reaction; primer; amplify; PCR; long; short;  
KW bacteriophage M13; gene III; E. coli; tryptophan synthase;  
KW beta-subunit; fusion sequence; ligand binding protein; antibody; ss.  
OS Synthetic.  
PN EP-614989-A.  
PD 14-SEP-1994.  
PD 16-FEB-1994.  
PF 17-FEB-1993; EP-102484.  
PR (MORP-) MORPHOSTS GES. PROTEINOPTIMERUNG MBH.  
PI Krebber C, Moroney S, Plueckthun A, Schneider C;  
DR WPI; 94-281214/35.  
PT Selection of genes encoding ligand binding proteins - using  
PT reproducible genetic packages infective only in presence of binding  
PT protein having high affinity for target ligand  
PS Example 1; Page 7; 26pp; English.  
CC The sequences given in Q65012-15 are primers which were used to  
CC amplify the long and short versions of the bacteriophage M13 gene  
CC III. The amplified sequence was ligated to the sequence of the E. coli  
CC by the primer sequences given in Q65016-19, which encodes the E. coli  
CC tryptophan synthase beta-subunit. When the fusion sequence is  
CC expressed it may be used to isolate ligand binding proteins, eg.  
CC antibodies which have a high affinity for a specific ligand, esp.  
CC from a large collection of variant molecules.  
SQ Sequence 45 BP; 9 A; 7 C; 13 G; 16 T;

Query Match	56.98;	Score 25.6;	DB 1;	Length 45;
Best Local Similarly	87.58;	Pred. No. 0.51;		
Matches 28; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0

Oy 4 aaagcttagtgcgtgcatgtaaaagcc 35  
|||  
Db 3 AAGCTTAGTGCATGTCGTGATTTCC 34

RESULT	4
ID	Q69013 standard; DNA; 42 BP
AC	Q69013;
DT	11-APR-1995 (first entry)

DE Bacteriophage M13 Gene III primer #2.  
 KW Polymerase chain reaction; primer; amplify; PCR; long; short;  
 KW Bacteriophage M13; gene III; *E. coli*; tryptophan synthase;  
 KW beta-subunit; fusion sequence; ligand binding protein; antibody; ss  
 OS Synthetic.

PA (MORP-) MORPHOSYS GES PROTEINOPTIMIERUNG MBH.  
PI Kriebler C, Moroney S, Pleuecthun A, Schneider C;  
DR WPr: 94-281214/35.  
PT Selection of genes encoding ligand binding proteins - using  
PT replicable genetic packages infective only in presence of binding  
PT protein having high affinity for target ligand  
PT Example 1; Page 7; 20pp; English.

CC The sequences given in 069015-15 are primers which were used to  
CC amplify the long and short versions of the bacteriophage M13 gene  
CC III. The amplified sequence was ligated to the sequence amplified  
CC by the primer sequences given in 069016-19, which encodes the E. coli  
CC tryptophan synthase beta-subunit. When the fusion sequence is  
CC expressed it may be used to isolate ligand binding proteins, eg.  
CC antibodies which have a high affinity for a specific ligand, esp.  
CC from a large collection of variant molecules.  
SC Sequence 42 BP; 11 A; 7 C; 15 G; 9 T;

Query Match	55.6%;	Score 25;	DB 1;	Length 42;
Best Local Similarity	100.0%;	Pred. No. 0.83;		
Matches 25; Conservative	0;	Mismatches	0;	Gaps 0;

QY	4	aagcttagtgatggtgatg	28
Db	3	AAGCTTAGTGATGCTGATG	27

RESULT 5  
069028  
069028 standard; DNA; 54 BP.  
ID 069028:  
AC 069028:  
DT  
DE 11-APR-1995 (first entry)  
DE Bacteriophage M13 Gene III primer.  
KW Polymerase chain reaction; primer; amplify; PCR; long; short;  
KW Bacteriophage M13; gene III; E. coli; cryptophan synthase;  
KW beta-subunit; fusion sequence; ligand binding protein; antibody; ss

PD 14-SEP-1994.  
PF 16-FEB-1994; 102334.  
PR 17-FEB-1993; EP-102464.  
PA (MOR-) MORPHOYS GES PROTEINOPTIMIERUNG MBH.  
PI Kriebder C, Moroney S, Plueckthun A, Schneider C;  
DR WPI: 94-781214/35.  
PT Selection of genes encoding ligand binding proteins - using  
PT replicable genetic packages infective only in presence of binding  
PT protein having high affinity for target ligand  
SS Example 1; Page 7; 20pp; English.

CC This sequence represents a primer which was used with the sequences  
CC given in 069012-12 to amplify the long and short versions of the  
CC bacteriophage M13 gene III. The amplified sequence was ligated to the  
CC sequence amplified by the primer sequences given in 069016-19, which  
CC encodes the E. coli tryptophan synthase beta-subunit. When the fusion  
CC sequence is expressed it may be used to isolate ligand binding proteins,  
CC eg. antibodies which have a high affinity for a specific ligand, esp.  
CC from a large collection of variant molecules. 54 BP; 14 A; 6 C; 21 G; 13 T;

Query Match	55.6%	Score 25;	DB 1;	Length 54;
Best Local Similarity	100.0%	Pred. No. 0.87;		
Matches 25; Conservative	0;	Mismatches	0;	Indels 0
				Gaps 0

QY 4 aagcttagtgatggtgatggtgatg 28







Query Match 51.1%; Score 23; DB 1; Length 51;  
 Best Local Similarity 83.9%; Pred. No. 4.5;  
 Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 8 ttagtgatgtgtatgtatgaaagccaa 38  
 |||  
 DB 48 TTCGTGATGTGATGTGATGAATTCGAA 18

# RESULT 15

VS1820  
 ID VS1820 standard; DNA; 58 BP.  
 AC VS1820;  
 DF 04-MAR-1999 (first entry)  
 DE Granulocytic Ehrlichia protein clone E46 PCR reverse primer.  
 KM Granulocytic ehrlichiosis; Ehrlichia sp.; GE protein; infection; tick;  
 diagnosis; vaccine; antigenic protein; antibody; immune response;  
 KM PCR primer; ss.  
 OS Synthetic.  
 OS Ehrlichia sp.  
 PN W09849313-A2.  
 PD 05-NOV-1998.  
 PE 24-APR-1998; U08265.  
 PR 25-APR-1997; US-044933.  
 PA (AQU1-) AQUILA BIOPHARMACEUTICALS INC.  
 PI Beltz GA, Coughlin RT, Murphy CA, Storey J;  
 DR WPI; 99-009432/01.  
 PT New nucleic acid from the human granulocytic ehrlichiosis agent -  
 PT and related antigenic proteins; vectors, transformed cells and  
 antibodies; useful for diagnosis and in protective vaccines  
 PS Example 15; Page 71; 154pp; English.  
 CC The present sequence represents a PCR primer for granulocytic ehrlichia  
 (GE) proteins used in an example from the present invention.  
 CC GE nucleic acids, vectors and host cells are used for the recombinant  
 production of GE proteins, and also in research to further characterise  
 the proteins. GE protein-encoding nucleic acid molecules are detected  
 by hybridisation to GE nucleic acid fragments or by using the fragments  
 as primers for polymerase chain reaction (PCR) amplification. GE  
 CC proteins, their immunogenic fragments, and GE nucleic acid molecules  
 encoding them are used to generate an immune response against GE,  
 CC specifically as (genetic) vaccines, especially to control ehrlichiosis  
 CC in humans and dogs, but also to raise Ab and to study DNA-protein  
 CC interactions. Ab are used to detect GE proteins by forming an immune  
 CC complex in standard assays, and correspondingly GE proteins can detect  
 CC specific antibodies, especially for diagnosis, assessment and prognosis  
 CC of GE infection, or of contamination of biological samples with GE.  
 SQ Sequence 58 BP; 15 A; 15 C; 17 G; 11 T;

Query Match 50.2%; Score 22.6; DB 1; Length 58;  
 Best Local Similarity 86.2%; Pred. No. 6.4;  
 Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 9 tagtgatgtgtatgtatgaaagccaa 37  
 |||  
 DB 12 TAGTGTGTGTGTGTGTGAAGCAAA 40

Search completed: June 4, 2000, 16:24:18  
 Job time: 28914 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 08:20:48 ; Search time 1236.38 Seconds  
(without alignments)

-13.376 Million cell updates/sec

Title: US-09-164-714-2

Perfect score: 17

Sequence: 1 gayencarcenctaygt 17

Scoring table: IDENTITY\_NUC

Searched: Gapop 10.0 , Gapext 1.0

Total number of hits satisfying chosen parameters: 356616

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : GenBml:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pr1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_v1:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_or:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl1:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_v1:\*  
32: gb\_htg1:\*  
33: gb\_htg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pr4:\*  
41: gb\_htg3:\*  
42: gb\_htg4:\*  
43: gb\_htg5:\*  
44: gb\_htg6:\*

45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pr5:\*  
52: gb\_htg8:\*  
53: gb\_htg9:\*  
54: gb\_htg10:\*  
55: gb\_htg11:\*  
56: gb\_htg12:\*  
57: gb\_htg13:\*  
58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	11.2	65.9	30	34	DRORN	K01289 D.melanog
2	11.2	65.9	52	14	SYNANVA4	M60112 Avian neov
3	11.2	65.9	76	16	AB003918	AB003918 Hepatitis
4	11.2	65.9	90	12	RATVPN	K02434 Rat vasopre
5	11.2	65.9	90	12	RATVPNP	K01495 Rat prepro
6	11.2	64.7	21	5	AR030742	AR030742 Sequence
7	11.2	64.7	21	5	AR030743	AR030743 Sequence
8	11.2	64.7	75	5	I25131	I25131 Sequence 21
9	11.2	64.7	90	5	I25143	I25143 Sequence 39
10	10.6	62.4	20	5	AR012271	AR012271 Sequence
11	10.6	62.4	20	5	I14969	I14969 Sequence 55
12	10.6	62.4	20	5	I73689	I73689 Sequence 55
13	10.6	62.4	35	5	I19619	I19619 Sequence 8
14	10.6	62.4	71	5	AR017688	AR017688 Sequence
15	10.2	60.0	21	5	A16125	A16125 Oligonucleo
16	10.2	60.0	21	5	A16126	A16126 Oligonucleo
17	10.2	60.0	21	5	I67007	I67007 Sequence 3
18	10.2	60.0	21	5	I67008	I67008 Sequence 4
19	10.2	60.0	30	5	A13425	A13425 Oligonucleo
20	10.2	60.0	30	5	A13426	A13426 Oligonucleo
21	10.2	60.0	30	5	A13427	A13427 Oligonucleo
22	10.2	60.0	30	5	A13428	A13428 Oligonucleo
23	10.2	60.0	30	5	A13541	A13541 Oligonucleo
24	10.2	60.0	30	5	A13542	A13542 Oligonucleo
25	10.2	60.0	30	5	A13543	A13543 Oligonucleo
26	10.2	60.0	30	5	A13544	A13544 Oligonucleo
27	10.2	60.0	30	5	A13545	A13545 Oligonucleo
28	10.2	60.0	30	5	A13546	A13546 Oligonucleo
29	10.2	60.0	30	5	A13547	A13547 Oligonucleo
30	10.2	60.0	30	5	A13548	A13548 Oligonucleo
31	10.2	60.0	30	5	A13568	A13568 Oligonucleo
32	10.2	60.0	30	5	A13569	A13569 Oligonucleo
33	10.2	60.0	30	5	I07186	I07186 Sequence 9
34	10.2	60.0	30	5	I30350	I30350 Sequence 4
35	10.2	60.0	32	5	A69250	A69250 Sequence 8
36	10.2	60.0	32	5	A71927	A71927 Sequence 8
37	10.2	60.0	32	5	E16812	E16812 PCR primer
38	10.2	60.0	32	5	E16831	E16831 PCR primer
39	10.2	60.0	35	5	A13549	A13549 Oligonucleo
40	10.2	60.0	35	5	A13550	A13550 Oligonucleo
41	10.2	60.0	36	5	I91771	I91771 Sequence 5
42	10.2	60.0	47	7	MIRGC13	MIRGC13 yeast mtoc
43	10.2	60.0	60	5	A21412	A21412 Oligonucleo
44	10.2	60.0	60	5	A21415	A21415 Oligonucleo
45	10.2	60.0	60	5	A21420	A21420 Oligonucleo

#### ALIGNMENTS





TITLE Vasopressin gene is expressed at low levels in the hypothalamus of the Brattleboro rat  
JOURNAL Proc Natl. Acad. Sci. U.S.A. 81, 5296-5299 (1984)  
MEDLINE 84296125  
COMMENT Over the region sequenced (1-53), the Brattleboro gene appears to be identical to the Long-Evans gene [1].  
FEATURES  
source 1..90  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
sig\_peptide 14..70  
/note="vasopressin-neurophysin signal pept"  
14..>90  
/note="vasopressin-neurophysin prepropeptide"  
/codon\_start=1  
/protein\_id="AAA42343.1"  
/db\_xref="GI:207676"  
/translation="MLNTTSLACFLSLALTSACYPQNC"  
mat\_peptide 71..>90  
/note="vasopressin"  
BASE COUNT 13 a 36 c 16 g 25 t  
ORIGIN unreported.

Query Match 65.9%; Score 11.2; DB 12; Length 90;  
Best Local Similarity 62.5%; Pred. No. 1.4e+04;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaycncarcntaygt 16  
||:| | | | | | | | | |  
Db 13 GATGCTCACTACG 28

RESULT 5  
LOCUS RATVPNP 90 bp mRNA ROD 27-APR-1993  
DEFINITION Rat preprovasopressin-neurophysin gene, 5' end.  
ACCESSION K01485  
VERSION K01485.1 GI:207677  
KEYWORDS neurophysin; vasopressin.  
SOURCE rat (Long-Evans) adult female hypothalamus, cDNA to mRNA, clones cDNA-1 and cDNA-2.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 90)  
AUTHORS Majzoub,J.A., Rich,A., van Boom,J.H. and Habener,J.F.  
TITLE Vasopressin and oxytocin mRNA regulation in the rat assessed by hybridization with synthetic oligonucleotides  
JOURNAL J. Biol. Chem. 258, 14061-14064 (1983)  
MEDLINE 84061831  
COMMENT When rats were given 2% saline solution to drink for 3 weeks the vasopressin levels in the hypothalamus increased 20-fold. There is one other entry under this title.  
FEATURES  
source 1..90  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
mrna <1..>90  
sig\_peptide 14..70  
/note="vasopressin-neurophysin mRNA"  
14..>90  
/note="preprovasopressin-neurophysin"  
/codon\_start=1  
/protein\_id="AAA42344.1"  
/db\_xref="GI:207678"  
/translation="MLNTTSLACFLSLALTSACYPQNC"  
mat\_peptide 71..>90  
/note="vasopressin"  
BASE COUNT 13 a 36 c 16 g 25 t  
ORIGIN

Query Match 65.9%; Score 11.2; DB 12; Length 90;  
Best Local Similarity 62.5%; Pred. No. 1.4e+04;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaycncarcntaygt 16  
||:| | | | | | | | | |  
Db 13 GATGCTCACTACG 28

RESULT 6  
LOCUS AR030742/c 21 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 28 from patent US 5861298.  
ACCESSION AR030742  
VERSION AR030742.1 GI:5943956  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 21)  
AUTHORS Adams,M.D., Blake,J.A., Debouck,C.M., Drake,F.H., Fitzgerald,L.M., Fraser,C.M., Gowen,M., Hastings,G.A., Kirkness,E.F., Lee,N.H. and Rood,J.  
TITLE Cathepsin K gene  
JOURNAL Patent: US 5861298-A 28 19-JAN-1999;  
FEATURES Location/Qualifiers  
source 1..21  
/organism="unknown"

BASE COUNT 5 a 5 c 5 g 6 t

Query Match 64.7%; Score 11; DB 5; Length 21;  
Best Local Similarity 64.7%; Pred. No. 1.9e+04;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gaycncarcntaygt 17  
||:| | | | | | | | | |  
Db 18 GATGCTACCATATG 2

RESULT 7  
LOCUS AR030743 21 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 29 from patent US 5861298.  
ACCESSION AR030743  
VERSION AR030743.1 GI:5943957  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 21)  
AUTHORS Adams,M.D., Blake,J.A., Debouck,C.M., Drake,F.H., Fitzgerald,L.M., Fraser,C.M., Gowen,M., Hastings,G.A., Kirkness,E.F., Lee,N.H. and Rood,J.  
TITLE Cathepsin K gene  
JOURNAL Patent: US 5861298-A 29 19-JAN-1999;  
FEATURES Location/Qualifiers  
source 1..21  
/organism="unknown"

BASE COUNT 6 a 5 c 5 g 5 t

Query Match 64.7%; Score 11; DB 5; Length 21;  
Best Local Similarity 64.7%; Pred. No. 1.9e+04;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gaycncarcntaygt 17  
||:| | | | | | | | | |  
Db 4 GATGCTACCATATG 20

RESULT 8  
LOCUS 125131/c 75 bp DNA PAT 07-OCT-1996  
DEFINITION Sequence 21 from patent US 5547871.  
ACCESSION 125131  
VERSION 125131.1 GI:1605001  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 75)  
AUTHORS Black,B.C. and Summers,M.D.  
TITLE Heterologous signal sequences for secretion of insect controlling proteins  
JOURNAL Patent: US 5547871-A 21 20-AUG-1996;  
FEATURES Location/Qualifiers  
source 1..75  
BASE COUNT 10 a 22 c 24 g 19 t  
ORIGIN

Query Match 64.7%; Score 11; DB 5; Length 75;  
Best Local Similarity 64.7%; Pred. No. 1.9e+04;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 gaygcncarcntaygt 17  
Db 36 GAGGCCACGCCGACCT 20

RESULT 9  
LOCUS 125143/c 90 bp DNA PAT 07-OCT-1996  
DEFINITION Sequence 39 from patent US 5547871.  
ACCESSION 125143  
VERSION 125143.1 GI:1605013  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 90)  
AUTHORS Black,B.C. and Summers,M.D.  
TITLE Heterologous signal sequences for secretion of insect controlling proteins  
JOURNAL Patent: US 5547871-A 39 20-AUG-1996;  
FEATURES Location/Qualifiers  
source 1..90  
BASE COUNT 17 a 25 c 28 g 20 t  
ORIGIN

Query Match 64.7%; Score 11; DB 5; Length 90;  
Best Local Similarity 64.7%; Pred. No. 1.9e+04;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 gaygcncarcntaygt 17  
Db 36 GAGGCCACGCCGACCT 20

RESULT 10  
LOCUS AR012271 20 bp DNA PAT 04-DEC-1998  
DEFINITION Sequence 55 from patent US 5763254.  
ACCESSION AR012271  
VERSION AR012271.1 GI:3970261  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 20)  
AUTHORS Woldike,H.Fabrizius, Hagen,F., Hjort,C.Malland and Hastrup,S.  
TITLE Enzyme capable of degrading cellulose or hemicellulose  
JOURNAL Patent: US 5686593-A 55 11-NOV-1997;  
FEATURES Location/Qualifiers  
source 1..20  
BASE COUNT 3 a 5 c 6 g 6 t  
ORIGIN

Query Match 62.4%; Score 10.6; DB 5; Length 20;  
Best Local Similarity 58.8%; Pred. No. 3.4e+04;  
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

REFERENCE 1 (bases 1 to 20)  
AUTHORS Woldike,H.Fabrizius, Hagen,F., Hjort,C.Malland and Hastrup,S.  
TITLE Enzyme capable of degrading cellulose or hemicellulose  
JOURNAL Patent: US 5763254-A 55 09-JUN-1998;  
FEATURES Location/Qualifiers  
source 1..20  
BASE COUNT 3 a 5 c 6 g 6 t  
ORIGIN

Query Match 62.4%; Score 10.6; DB 5; Length 20;  
Best Local Similarity 58.8%; Pred. No. 3.4e+04;  
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 gaygcncarcntaygt 17  
Db 3 GATGCTCAGTCTACCT 19

RESULT 11  
LOCUS 114969 20 bp DNA PAT 02-APR-1996  
DEFINITION Sequence 55 from patent US 5457046.  
ACCESSION 114969  
VERSION 114969.1 GI:1249877  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 20)  
AUTHORS Woldike,H.F., Hagen,F., Hjort,C. and Hastrup,S.  
TITLE Enzyme capable of degrading cellulose or hemicellulose  
JOURNAL Patent: US 5457046-A 55 10-OCT-1995;  
FEATURES Location/Qualifiers  
source 1..20  
BASE COUNT 3 a 5 c 6 g 6 t  
ORIGIN

Query Match 62.4%; Score 10.6; DB 5; Length 20;  
Best Local Similarity 58.8%; Pred. No. 3.4e+04;  
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 gaygcncarcntaygt 17  
Db 3 GATGCTCAGTCTACCT 19

RESULT 12  
LOCUS 173689 20 bp DNA PAT 03-APR-1998  
DEFINITION Sequence 55 from patent US 5686593.  
ACCESSION 173689  
VERSION 173689.1 GI:3009830  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 20)  
AUTHORS Woldike,H.Fabrizius, Hagen,F., Hjort,C.Malland and Hastrup,S.  
TITLE Enzyme capable of degrading cellulose or hemicellulose  
JOURNAL Patent: US 5686593-A 55 11-NOV-1997;  
FEATURES Location/Qualifiers  
source 1..20  
BASE COUNT 3 a 5 c 6 g 6 t  
ORIGIN

Query Match 62.4%; Score 10.6; DB 5; Length 20;  
Best Local Similarity 58.8%; Pred. No. 3.4e+04;  
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 gaygncarccontaygt 17  
||:| | | | | | | | | |  
Db 3 GATGCTCAGTGTACGT 19

RESULT 13  
119619  
LOCUS 119619 35 bp DNA PAT 07-OCT-1996  
DEFINITION Sequence 8 from patent US 5510099.  
ACCESSION 119619  
VERSION 119619.1 GI:1599974  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 35)  
AUTHORS Short,J.M. and Kretz,P.L.  
TITLE Mutagenesis testing using transgenic non-human animals carrying  
test DNA sequences  
JOURNAL Patent: US 5510099-A 8 23-APR-1996;  
FEATURES Location/Qualifiers  
source 1..35  
/organism="unknown"  
BASE COUNT 9 a 11 c 8 g 7 t  
ORIGIN

Query Match 62.4%; Score 10.6; DB 5; Length 35;  
Best Local Similarity 58.8%; Pred. No. 3.4e+04;  
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 gaygncarccontaygt 17  
||:| | | | | | | | | |  
Db 10 GACGATCAGCCAAACGT 26

RESULT 14  
AR017688 71 bp DNA PAT 05-DEC-1998  
LOCUS AR017688  
DEFINITION Sequence 89 from patent US 5780228.  
ACCESSION AR017688  
VERSION AR017688.1 GI:3973291  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 71)  
AUTHORS Parma,D.H., Hicke,B., Bridonneau,P. and Gold,L.  
TITLE High affinity nucleic acid ligands to lectins  
JOURNAL Patent: US 5780228-A 89 14-JUL-1998;  
FEATURES Location/Qualifiers  
source 1..71  
/organism="unknown"  
BASE COUNT 19 a 15 c 26 g 11 t  
ORIGIN

Query Match 62.4%; Score 10.6; DB 5; Length 71;  
Best Local Similarity 58.8%; Pred. No. 3.4e+04;  
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 gaygncarccontaygt 17  
||:| | | | | | | | | |  
Db 34 GACCCGCTACTTATCT 18

RESULT 15  
A16125 21 bp DNA PAT 03-OCT-1994  
LOCUS A16125/C  
DEFINITION Oligonucleotide adaptor (SEQ ID NO: 8).  
ACCESSION A16125

VERSION A16125.1 GI:640908  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
REFERENCE 1 (bases 1 to 21)  
AUTHORS Cornelissen,B.J.C., Melchers,L.S., Meulenhoff,E.J.S., Van  
Roekel,J.S.C., Sela-Buurlage,M.B., Vloemans,A.A., Woloshuk,C.P.,  
Bol,J.F. and Linthorst,H.J.M.  
TITLE Fungal resistant plants, process for obtaining fungal resistant  
plants and recombinant polynucleotides for use therein  
JOURNAL Patent: EP 0440304-A 8 07-AUG-1991;  
FEATURES MOGEN INTERNATIONAL N.V.; Rijksuniversiteit te Leiden  
Location/Qualifiers  
source 1..21  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 6 a 5 c 6 g 4 t  
ORIGIN

Query Match 60.0%; Score 10.2; DB 5; Length 21;  
Best Local Similarity 60.0%; Pred. No. 6.2e+04;  
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 ygcncarccontaygt 17  
:| | | | | | | | | |  
Db 21 TGCACAGCCTCATGT 7

Search completed: June 4, 2000, 16:05:23  
Job time: 27875 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 08:22:24 ; Search time 322.35 seconds

(without alignments)  
13.195 Million cell updates/sec

Title: US-09-164-714-2

Perfect score: 17

Sequence: 1 gaycncarcntaygt 17

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 431286

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	11	64.7	21	1 V09686	Human cathepsin K
C 2	11	64.7	21	1 V09687	Human cathepsin K
C 3	11	64.7	29	1 06652	PCR primer P1 to a
C 4	11	64.7	59	1 050214	c-fos position 272
C 5	11	64.7	63	1 080732	Drosophila melanog
C 6	11	64.7	63	1 V18214	Codon optimised si
C 7	11	64.7	75	1 067694	Codon-optimised D.
C 8	11	64.7	90	1 067706	Fragment B esteras
C 9	11	64.7	90	1 067329	AAT optimised seq
C 10	11	64.7	90	1 V18322	Oligomer for a cod
C 11	10.6	62.4	19	1 V72009	Electronic perturb
C 12	10.6	62.4	24	1 V11542	Recombinant MPO DN
C 13	10.6	62.4	26	1 V39432	Humanised anti-HM1
C 14	10.6	62.4	35	1 048146	Lambda 37225-37255
C 15	10.6	62.4	35	1 V39428	Humanised anti-HM1
C 16	10.6	62.4	38	1 V10832	Human MSH2 gene PC
C 17	10.6	62.4	38	1 V39427	Humanised anti-HM1
C 18	10.6	62.4	50	1 V39429	Humanised anti-HM1
C 19	10.6	62.4	50	1 V39430	Humanised anti-HM1
C 20	10.6	62.4	71	1 T57762	L-selectin family
C 21	10.6	62.4	75	1 011149	Probe GTR-1 based
C 22	10.2	60.0	18	1 T50625	Human CERP hairpin
C 23	10.2	60.0	20	1 X15745	PCR primer used to
C 24	10.2	60.0	30	1 N90827	Anti-R gene constr
C 25	10.2	60.0	30	1 N95064	HTLV-1 R homologue
C 26	10.2	60.0	30	1 068400	HTLV-1 R gene c
C 27	10.2	60.0	30	1 T47837	HTLV-1 R region ol
C 28	10.2	60.0	30	1 T72860	Alpha-1-antitryps
C 29	10.2	60.0	32	1 V49438	Primer AB028 for B
C 30	10.2	60.0	32	1 V53672	Primer for T. pall
C 31	10.2	60.0	36	1 097842	Human E2 CDNA prim
C 32	10.2	60.0	36	1 097839	Human E1 CDNA prim
C 33	10.2	60.0	36	1 T41762	Ubiquitin-actinall
C 34	10.2	60.0	40	1 072625	Primer P4, for sim

C 35	10.2	60.0	40	1 072627	Primer P6, for sim
C 36	10.2	60.0	50	1 T69252	Flt-1 gene probe B
C 37	10.2	60.0	60	1 010918	Tag DNA polymerase
C 38	10.2	60.0	61	1 010919	Tag DNA polymerase
C 39	10.2	60.0	63	1 010917	Tag DNA polymerase
C 40	10.2	60.0	63	1 010916	Tag DNA polymerase
C 41	10.2	60.0	63	1 010925	Tag DNA polymerase
C 42	10.2	60.0	63	1 057919	Diversified HIV-1
C 43	10.2	60.0	67	1 V82651	Oligonucleotide us
C 44	10.2	60.0	92	1 X23496	Human neutrophil c
C 45	10.2	60.0	97	1 040278	Sequence of synthe

## ALIGNMENTS

```
RESULT 1
ID V09686 standard; DNA; 21 BP.
AC V09686:
DF 20-JUL-1998 (first entry)
DE Human cathepsin K gene exon 5 reverse PCR primer 4R.
KW Cathepsin K; human; osteoporosis; periodontal disease;
KW Paget's disease; Gaucher's disease; Alzheimer's disease;
KW central nervous system inflammation; hyperparathyroidism;
KW bone degradation; dental implant degradation; metastasis; tumour;
KW diagnosis; therapy; marker; PCR; primer; ss.
OS Synthetic.
OS Homo sapiens.
PN EP-812916-A2.
PD 17-DEC-1997.
PF 19-MAY-1997; 303395.
PR 26-AUG-1996; US-026083.
PR 14-JUN-1996; US-019942.
PR 17-JUN-1996; US-020273.
PR 26-AUG-1996; WO-014026.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (GENO-) INST GENOMIC RES.
PA (SMIR-) SMITHKLINE BEECHAM CORP.
PI Adams MD, Blake JA, Debonck CM, Drake FH, Fitzgerald LM,
PI Fraser CM, Gowen M, Hastings GA, Kirkness EF, Lee NH,
PI Road J;
DR WPI: 98-034977/04.
PT DNA encoding human cathepsin K - useful for diagnosing and treating
PT diseases associated with cathepsin K e.g. osteoporosis, bone
PT degradation, metastatic tumours, etc
PT Example 1; Page 50; 84pp; English.
CC CC (see V09670) of the human cathepsin K gene (see V09660). PCR
CC primers (see V09679-90) to adjacent exons of the cathepsin K
CC gene were used in the amplification of human genomic DNA. DNA
CC sequencing of intron-exon boundaries allowed sequencing of the
CC cathepsin genomic DNA. DNA encoding human cathepsin K is useful
CC for the diagnosis and treatment of e.g. osteoporosis, periodontal
CC disease, Paget's disease, Gaucher's disease, CNS inflammation,
CC Alzheimer's disease, hyperparathyroidism, bone degradation,
CC metastatic tumours, and degradation of bone implants and
CC prostheses, especially dental implants.
SQ Sequence 21 BP; 5 A; 5 C; 5 G; 6 T;

Query Match 64.7%; Score 11; DB 1; Length 21;
Best Local Similarity 64.7%; Pred. No. 4.3e+02;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 gaycncarcntaygt 17
Db 18 GATCCTACCAATAGT 2

RESULT 2
V09687
ID V09687 standard; DNA; 21 BP.
```

AC	V09687;	OS	V09687;	OS	Synthetic.	KW	Human cathepsin K gene exon 5, 6 forward PCR primer 5+6F.	DR	WPI: 98-034977/04.	PT	DNA encoding human cathepsin K - useful for diagnosing and treating	CC	primers (see V09679-90) to adjacent exons of the cathepsin K	CC	sequencing of intron-exon boundaries allowed sequencing of the	CC	cathepsin genomic DNA. DNA encoding human cathepsin K is useful	CC	for the diagnosis and treatment of e.g. osteoporosis, periodontal	CC	disease, Paget's disease, Gaucher's disease, CNS inflammation,	CC	Alzheimer's disease, hyperparathyroidism, bone degradation,	CC	metastatic tumours, and degradation of bone implants and	CC	prostheses, especially dental implants.	SO	Sequence 21 bp; 6 A; 5 C; 5 G; 5 T;	QY	1 gaygencarcantaygt 17	DB	4 GATGCTTACCATATGT 20	RESULT 3	066652/c	ID	Q66652 standard; DNA: 29 BP.	AC	Q66652;	DT	19-JAN-1995 (first entry)	DE	PCR primer p1 to amplify human dystrophin gene.	KW	Polymerase chain reaction amplification; human dystrophin gene; PCR;	KW	ss.	OS	Synthetic.	PN	S01792972-A.	PD	07-FEB-1993.	PR	15-MAY-1990; 826213.	PR	15-MAY-1990; SU-826213.	PA	(MOL=) MOL DIAGNOSIS METHODS RES CENTRE.	PI	Evgrafov OV, Polyakov AV;	PT	WPI: 94-181019/22.	PT	Amplification of DNA fragment of unknown sequence - giving wider	PT	functional possibilities because of increased distance between	PT	DNA fragment of known sequence and multiplied DNA fragment of	PT	unknown sequence	SS	Example; Column 4; 4bp; Russian.	CC	Two oligonucleotide primers p1 and p2 (Q66652 and Q66653,
----	---------	----	---------	----	------------	----	---	----	--------------------	----	---	----	--	----	--	----	---	----	---	----	--	----	---	----	--	----	---	----	-------------------------------------	----	------------------------	----	-----------------------	----------	----------	----	------------------------------	----	---------	----	---------------------------	----	---	----	--	----	-----	----	------------	----	--------------	----	--------------	----	----------------------	----	-------------------------	----	--	----	---------------------------	----	--------------------	----	--	----	--	----	---	----	------------------	----	----------------------------------	----	---

Query Match	Best Local Similarity	Matches	Score 11:	DB 1:	Length 50:
1 gaygncarcctatggt 17	64.7%	Score 11: DB 1: Length 50:	64.7%	Pred. No. 5.5e+02:	0;
11:	64.7%	2: Mismatches 4; Indels 0; Gaps 0;			
19 GAAGGACAGCCATTACGT 35					

```
RESULT 5
080732/c
ID 080732 standard; DNA; 63 BP.
AC 080732;
DT 22-AUG-1995 (first entry)
DE Drosophila melanogaster esterase-6 signal peptide gene.
KM Esterase-6 signal peptide gene; Drosophila melanogaster;
KW modular expression vectors; ss.
OS Drosophila melanogaster.
PN W09428114-A.
PD 08-DEC-1994.
PF 27-MAY-1994; U06079.
PR 28-MAY-1993; US-070164.
PA (AMCY ) AMERICAN CYANAMID CO.
PI Dierks PM, Webb NR;
DR WPI; 95-02278/03.
PT Recombinant double stranded DNA insect virus - comprising a
PT modular expression vector and a direct ligation virus vector
PS Disclosure; Page 85; 114pp; English.
CC 080732 encodes the native esterase-6 signal peptide from Drosophila
CC melanogaster. This sequence was used in the construction of a modular
CC expression vector, which was combined with a direct ligation virus
CC vector to produce a recombinant ds DNA insect virus. The ability to
CC directly insert foreign DNA into the virus is advantageous, as no
CC transfer or plasmid vectors are required.
SQ Sequence 63 BP; 9 A; 14 C; 22 G; 18 T;

Query Match 64.7%; Score 11; DB 1; Length 63;
Best Local Similarity 64.7%; Pred. No. 5.7e+02;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 gaycncarcntaygt 17
DB 24 GAGGCCCGAGCCGACGT 8

RESULT 6
V18214/c
ID V18214 standard; DNA; 63 BP.
AC V18214;
DT 02-SEP-1998 (first entry)
DE Codon optimised signal sequence of esterase-6.
KW Codon optimised; North African scorpion; Insect toxin; AAT;
KW Insect cell; expression; plant protection; Insect damage;
KW Insect virus; nuclear polyhedrosis virus; granulosis virus;
KW non-occluded virus; entomopox virus; ss.
OS Synthetic.
PN A09748372-A.
PD 12-MAR-1998.
PF 15-DEC-1997; 048372.
PR 25-JAN-1993; US-009264.
PA (AMCY ) AMERICAN CYANAMID CO.
PI Black BC, Brennan LA, Dierks PM;
DR WPI; 98-241214/22.
PT Isolated; codon optimised nucleic acid sequence encoding Androctonus
PT australis Insect toxin - useful for protecting plants from insect
PT damage by delivering insect virus containing it
PS Disclosure; Page 22; 81pp; English.
CC V18209-15 represent codon optimised signal sequences used in the
CC course of the invention. The specification describes a codon optimised
CC sequence for the Androctonus australis (North African scorpion) insect
CC toxin (AAT) gene. A. australis binds to sodium channels in insects and
CC causes contractile paralysis in insect larvae. Codon usage of the wild
CC type gene (see V18208) is optimised for expression in insect cells (see
CC V18207). The sequence is combined with a secretory signal sequence, also
CC optimised for expression in insect cells (see V18209-15). The optimised
CC sequence is useful for producing recombinant AAT, especially in insect
CC cells, or for protecting plants from insect damage. The toxin gene is
CC delivered by an insect virus, such as nuclear polyhedrosis virus,
CC granulosis virus, non-occluded virus, or entomopox virus.
SQ Sequence 63 BP; 9 A; 14 C; 22 G; 18 T;
```

```
Query Match 64.7%; Score 11; DB 1; Length 63;
Best Local Similarity 64.7%; Pred. No. 5.7e+02;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 gaycncarcntaygt 17
DB 24 GAGGCCCGAGCCGACGT 8

RESULT 7
067694/c
ID 067694 standard; DNA; 75 BP.
AC 067694;
DT 13-MAR-1995 (first entry)
DE Codon-optimised D. melanogaster esterase-6 signal sequence DNA.
KW Codon optimised; native; B. mori; pMHPC-12; signal peptide; chorion;
KW Insect controlling protein; toxin; AAT; baculovirus; ACMPV; Cry IVD;
KW expression; secretion; toxin-induced paralysis; cuticle; apolipophorin;
KW sex-specific; adipokinetic; esterase-6; D. melanogaster; neuropeptide;
KW M. sexta; enzyme; pyrenoxen; B. thuringiensis; diuretic hormone;
KW eclosion hormone; prothoracicotropic hormone; adipokinetic hormone;
KW proctolin; juvenile hormone esterase; ss.
OS Drosophila melanogaster.
FH Key location/Qualifiers
FT cds 13..75
FT /tag= a
FT /product= Esterase-6 signal sequence

PN EP-608696-A.
PD 03-AUG-1994.
PF 10-JAN-1994; 100265.
PR 25-JAN-1993; US-009265.
PA (AMCY ) AMERICAN CYANAMID CO.
PI Black BC;
DR WPI; 94-242108/30.
DR P-PDB: R56466.
PT Heterologous signal sequences for secretion of insect controlling
PT proteins - useful to protect plants from insect pests
PS Disclosure; Page 38; 69pp; English.
CC The sequences given in 067694-95 represent the codon optimised and
CC native coding sequences for the D. melanogaster esterase-6 signal
CC peptide, respectively. This signal peptide sequence may be used with
CC a DNA sequence encoding an insect controlling protein, esp. the insect-
CC specific toxin AAT. The fusion sequence may be introduced into an
CC insect virus, such as the baculovirus ACMPV. The insertion of the
CC AAT gene and the heterologous signal sequence into a baculovirus
CC results in the expression and secretion of the toxin. A susceptible
CC insect which ingests such a modified baculovirus will cease feeding
CC on plants due to toxin-induced paralysis at an earlier time than an
CC insect which ingests a wild-type baculovirus, thus reducing crop
CC damage.
SQ Sequence 75 BP; 10 A; 22 C; 24 G; 19 T;

Query Match 64.7%; Score 11; DB 1; Length 75;
Best Local Similarity 64.7%; Pred. No. 5.8e+02;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 gaycncarcntaygt 17
DB 36 GAGGCCCGAGCCGACGT 20

RESULT 8
067706/c
ID 067706 standard; DNA; 90 BP.
AC 067706;
DT 13-MAR-1995 (first entry)
DE Fragment B esterase-6 oligomer; encodes esterase-6 signal sequence.
KW Codon optimised; native; B. mori; pMHPC-12; signal peptide; chorion;
KW Insect controlling protein; toxin; AAT; baculovirus; ACMPV; Cry IVD;
KW expression; secretion; toxin-induced paralysis; cuticle; apolipophorin;
KW sex-specific; adipokinetic; esterase-6; D. melanogaster; neuropeptide;
```





/note- "3'-end modified by presence of Biotin"

FT MO9851819-A1.  
 PN 19-NOV-1998.  
 PD 07-MAY-1998; 009357.  
 PF 14-MAY-1997; US-855058.  
 PA (NANO-) NANOGEN INC.  
 PI Heller MJ, O Connell JP, Sosnowski RG, Tu E;  
 DR MPI: 99-059702/05  
 PT Hybridisation analysis using electronic stringency control device  
 PT based on changes in fluorescence - when an electric field is applied  
 PT to the hybrid, used e.g. for sequencing or for detecting single  
 PT mismatch mutations, also electronic perturbation catalysis  
 PS Example 1: Page 28; 56pp; English.  
 CC This sequence represents a probe used in a method for studying  
 CC hybridisation analysis by detecting electronic denaturation using  
 CC electronic perturbation analysis. The method is particularly used for  
 CC sequencing and to discriminate between match and mismatch hybrids,  
 CC particularly to detect point mutations, deletions, inserts, repeat  
 CC regions, single nucleotide polymorphisms, translocations, intron/exon  
 CC junctions, etc. The same principle may be used for most molecular  
 CC biological processes, e.g. antibody-antigen reactions, cell typing and  
 CC separation, enzymatic and other clinical assays, also in optoelectronic  
 CC devices and optical memory materials. The method is applicable to any  
 CC type of reaction, e.g. nanofabrication or other self-assembling or  
 CC self-organising processes, synthesis of nucleic acids, peptides, polymers  
 CC etc. The method relies on the observation that perturbation of a  
 CC fluorescence signal (particularly a peak) occurs at the electric power  
 CC level which causes denaturation or dehybridisation. The method is very  
 CC rapid, e.g. match/mismatch hybrids can be distinguished in less than a  
 CC minute, e.g. 5 sec. It permits use of long probes (over 20 bases) and  
 CC probe specificity is determined by the position of the label. Since  
 CC analysis does not require removal of mismatched probes, the sample can be  
 CC analysed repeatedly to improve assay statistics.  
 SQ Sequence 19 BP; 4 A; 4 C; 7 G; 4 T;

Query Match 62.4%; Score 10.6; DB 1; Length 19;  
 Best Local Similarity 58.8%; Pred. No. 8.5e+02;  
 Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 gaycncarcantaygt 17  
 ||:| ||:| ||:| ||:|  
 DB 1 GATGACGAGCTTACGT 17

RESULT 12  
 ID V11542/c  
 ID V11542 standard; DNA; 24 BP.  
 AC V11542;  
 DT 30-JUL-1998 (first entry)  
 DE Recombinant MPO DNA H7b 3'-end primer.  
 KM Myeloperoxidase; MPO; MPO-ANCA; immunoreaction; IgG; diagnosis;  
 KM Myeloperoxidase-specific anti-neutrophil cytoplasmic antibody; primer;  
 KW rheumatoid arthritis; systemic lupus erythematosus; nephritis; ss.  
 OS Synthetic.  
 PN WO9807848-A1.  
 PD 26-FEB-1998.  
 PF 21-AUG-1997; J02910.  
 PR 21-AUG-1996; JP-219855.  
 PA (TEIKO) TEIKOKU HORMONE MFG CO LTD.  
 PI Suzuki K, Tanokura M;  
 DR MPI: 98-169154/15.  
 PT Recombinant myeloperoxidase (MPO) fragments reactive with MPO  
 PT auto-antibody - for use in diagnosis of auto-immune diseases by  
 PT auto-antibody assay in blood samples  
 PS Disclosure: Page 9; 45pp; Japanese.  
 CC Primers V11523-V11546 are used in a method which produces recombinant  
 CC myeloperoxidase (MPO) fragments that are used to assay MPO reactivity  
 CC with a myeloperoxidase-specific anti-neutrophil cytoplasmic antibody  
 CC (MPO-ANCA). The MPO fragments may be fused with a histidine hexamer,  
 CC with maltose-binding protein or with glutathione S-transferase (GST).  
 CC The MPO fragments may be used for assay of MPO-ANCA in biological  
 CC samples such as blood by an immunoreaction of the sample with immobilised

CC MPO fragment, then a further reaction with labelled anti-human IgG  
 CC antibody (e.g. using a horseradish peroxidase, alkali phosphatase or  
 CC beta-galactosidase label), then a final assay of the bound label (e.g.  
 CC by reaction with o-phenylenediamine for the peroxidase, or with  
 CC p-nitrophenyl phosphate for the phosphatase or galactosidase). This assay  
 CC of MPO-ANCA in biological samples can be used for the diagnosis of  
 CC autoimmune diseases such as rheumatoid arthritis, systemic lupus  
 CC erythematosus and autoimmune nephritis.  
 SQ Sequence 24 BP; 5 A; 5 C; 8 G; 6 T;

Query Match 62.4%; Score 10.6; DB 1; Length 24;  
 Best Local Similarity 58.8%; Pred. No. 8.8e+02;  
 Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 gaycncarcantaygt 17  
 ||:| ||:| ||:| ||:|  
 DB 17 GATGCCAGCTTACCT 1

RESULT 13  
 ID V39432  
 ID V39432 standard; DNA; 26 BP.  
 AC V39432;  
 DT 21-SEP-1998 (first entry)  
 DE Humanised anti-HM1.24 antibody PCR primer SEQ ID NO:93.  
 KM Mouse; human; humanised; anti-HM1.24 antibody; myeloma; FR; CDR;  
 KW framework region; complementarity determining region; antigenicity;  
 KM PCR primer; ss.  
 OS Synthetic.  
 OS Mus sp.  
 OS Homo sapiens.  
 PN WO9814580-A1.  
 PD 09-APR-1998.  
 PF 03-OCT-1997; J03553.  
 PR 04-OCT-1996; JP-264756.  
 PA (CHUS) CHUGAI SEIRYAKU KK.  
 PI Koshihara Y, Kosaka M, Ontomo T, Ono K, Tsuchiya M,  
 PI Yoshimura Y;  
 DR MPI: 98-286421/25.  
 PT Humanised anti-HM1.24 antibody - for treatment of myeloma  
 PS Example 9: Page 147; 210pp; Japanese.  
 CC A humanised anti-HM1.24 antibody has been developed which comprises  
 CC human L and H chain C regions, and L and/or H chain V regions  
 CC containing material originating in mouse anti-HM1.24 antibody. The V  
 CC regions contain framework (FR) regions of human origin and  
 CC complementarity determining regions (CDR) of mouse origin, leading to  
 CC a reshaped humanised antibody. The C regions are human Ck (L-chain) and  
 CC human C gamma (especially C gamma 1) (H-chain). The FR regions of the  
 CC L chain V region are derived from human subtype HSG1 (e.g. from human  
 CC antibody RE1) and the FR regions of the H chain V region are derived  
 CC from human subtype HSG1 (e.g. FR1-3 from human antibody HG3 and FR4  
 CC from human antibody JH6). The present sequence represents a PCR primer  
 CC used in an example from the present invention. The antibodies are used  
 CC for the treatment of myeloma, especially by injection, intravenously  
 CC intramuscularly or subcutaneously. The antibodies are used at 0.01-1000  
 CC (especially 5-100) mg/kg body weight. The humanised antibody has low  
 CC antigenicity and is therefore therapeutically in humans.  
 SQ Sequence 26 BP; 6 A; 7 C; 6 G; 7 T;

Query Match 62.4%; Score 10.6; DB 1; Length 26;  
 Best Local Similarity 58.8%; Pred. No. 8.9e+02;  
 Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 gaycncarcantaygt 17  
 ||:| ||:| ||:| ||:|  
 DB 9 GATCTCAGCTTCATGT 25

RESULT 14  
 ID Q48146  
 ID Q48146 standard; DNA; 35 BP.

```

AC 048146:
DE 10-FEB-1994 (first entry)
DE Lambda 37225-37255 primer p2.
KW primer; PCR; polymerase chain reaction; polysogenic; microorganism;
KW polysogen; E. coli; strain SCS-8; lambda; ci gene; SCS-8ci; puc18;
KW N99; putKm; putKm-ci; kanamycin; transformation; s17-1/putKm-ci;
KW puc18sfi1; conjugation; tetracycline; lytic phase; ss.
OS Synthetic.
PN WO9315769-A.
PD 19-AUG-1993.
PF 12-FEB-1993: U01293.
PR 14-FEB-1992; US-837031.
PR 04-SEP-1992; US-941190.
PA (STRA-) STRATAGENE.
PI Kretz PL, Short JM.
DR WPI: 93-272571/34.
PT Mutagenesis testing using transgenic non-human animals carrying
PT test DNA sequences - providing selectable reporter molecules for
PT high sensitivity and of detection
PR Disclosure; Page 90; 145pp; English.
CC The sequences given in 048145-46 are primers which were used in the
CC preparation of a polysogenic microorganism. The polysogen was
CC constructed in E. coli strain SCS-8 and a stably integrated copy of
CC the lambda ci gene was inserted into the genome to form the E. coli
CC designated SCS-8ci. The plasmid puc18sfi1 was derived from puc18 by
CC adding two SfiI restriction sites flanking the polylinker. These two
CC primers correspond to the termini of the wild type ci gene and were
CC used on wild type lambda in the form of an E. coli lysogen extract of
CC strain N99. The amplification product was ligated into the polylinker
CC region of the plasmid puc18sfi1 and then removed from the plasmid by
CC SfiI digestion, to form a PCR product with SfiI cohesive ends. putKm
CC was digested with SfiI and the PCR fragment was ligated into it to
CC form putKm-ci. putKm-ci was transformed into E. coli and cultured in
CC kanamycin to select for transformants containing the plasmid, and were
CC designated s17-1/putKm-ci. A standard mating protocol was conducted
CC where cultures of SCS-8 and s17-1/putKm-ci were coinoculated. Plasmid
CC putKm-ci was transferred by conjugation to SCS-8 cells and the
CC resulting conjugates were plated onto plates containing kanamycin and
CC tetracycline. Viable cells were designated SCS-8ci cells and are
CC polysogenic microorganisms as the expression of the ci gene prevents
CC the cells entering the lytic phase upon infection by a lysis-competent
CC lambda bacteriophage.
CC Sequence 35 BP; 9 A; 11 C; 8 G; 7 T;
SQ

```

Query Match 62.4%; Score 10.6; DB 1; Length 35;  
Best Local Similarity 58.8%; Pred. No. 9.2e+02;  
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

OY 1 gaygncarcrcntaygt 17
   ||:| ||:| ||:| ||:|
DB 10 GAGGATCAGCCAAACGT 26

```

RESULT 15  
V39428  
ID V39428 standard; DNA: 35 BP.  
AC V39428;  
DE 21-SEP-1998 (first entry)  
DE Humanised anti-HM1.24 antibody PCR primer SEQ ID NO:89.  
KW Mouse; human; humanised; anti-HM1.24 antibody; myeloma; FR; CDR;  
KW framework region; complementarity determining region; antigenicity;  
KW PCR primer; ss.  
OS Synthetic.  
OS Mus sp.  
OS Homo sapiens.  
PN WO9814580-A1.  
PD 09-APR-1998.  
PF 03-OCT-1997; J03553.  
PR 04-OCT-1996; JP-264756.  
PA (CHUS ) CHUGAI SEIYAKU KK.  
PI Kojishimura Y, Kosaka M, Ohtomo T, Ono K, Tsuchiya M,  
PI Yoshimura Y;

```

DR WPI: 98-286421/25.
PT Humanised anti-HM1.24 antibody - for treatment of myeloma
PS Example 9; Page 146; 210pp; Japanese.
CC A humanised anti-HM1.24 antibody has been developed which comprises
CC human L and H chain C regions, and L and/or H chain V regions
CC containing material originating in mouse anti-HM1.24 antibody. The V
CC regions contain framework (FR) regions of human origin and
CC complementarity determining regions (CDR) of mouse origin, leading to
CC a reshaped humanised antibody. The C regions are human Ck (L-chain) and
CC human C gamma (especially C gamma 1) (H-chain). The FR regions of the
CC L chain V region are derived from human subtype HSG1 (e.g. from human
CC antibody RE1) and the FR regions of the H chain V region are derived
CC from human subtype HSG1 (e.g. FR1-3 from human antibody HG3 and FR4
CC from human antibody JH6). The present sequence represents a PCR primer
CC used in an example from the present invention. The antibodies are used
CC for the treatment of myeloma, especially by injection. Intravenously,
CC intramuscularly or subcutaneously. The antibodies are used at 0.01-1000
CC (especially 5-100) mg/kg body weight. The humanised antibody has low
CC antigenicity and is therefore effective therapeutically in humans.
SQ Sequence 35 BP; 6 A; 8 C; 11 G; 10 T;

```

Query Match 62.4%; Score 10.6; DB 1; Length 35;  
Best Local Similarity 58.8%; Pred. No. 9.2e+02;  
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

OY 1 gaygncarcrcntaygt 17
   ||:| ||:| ||:| ||:|
DB 9 GATGCTCAGCTGCACT 25

```

Search completed: June 4, 2000, 16:23:57  
Job time: 28893 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 08:04:16 ; Search time 4521.53 Seconds  
(without alignments)  
15.239 Million cell updates/sec

Title: US-09-164-714-2

Perfect score: 17  
Sequence: 1 gaycncarcntaygt 17

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 119968

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: gb\_est1: \*  
21: gb\_est2: \*  
22: gb\_est3: \*  
23: gb\_est4: \*  
24: gb\_est5: \*  
25: gb\_est6: \*  
26: gb\_est7: \*  
27: gb\_est8: \*  
28: gb\_est9: \*  
29: gb\_est10: \*  
30: gb\_est11: \*  
31: gb\_est12: \*  
32: gb\_est13: \*  
33: gb\_est14: \*  
34: gb\_est15: \*  
35: gb\_est16: \*  
36: gb\_est17: \*  
37: gb\_est18: \*  
38: gb\_est19: \*  
39: gb\_est20: \*  
40: gb\_est21: \*  
41: gb\_est22: \*  
42: gb\_est23: \*  
43: gb\_est24: \*  
44: gb\_est25: \*

45: gb\_est26: \*  
46: gb\_est27: \*  
47: gb\_est28: \*  
48: gb\_est29: \*  
49: gb\_est30: \*  
50: gb\_est31: \*  
51: gb\_est32: \*  
52: em\_est20: \*  
53: em\_est21: \*  
54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*  
59: gb\_est33: \*  
60: gb\_est34: \*  
61: gb\_est35: \*  
62: gb\_est36: \*  
63: gb\_est37: \*  
64: gb\_est38: \*  
65: em\_est27: \*  
66: em\_est28: \*  
67: em\_est29: \*  
68: em\_est30: \*  
69: gb\_est39: \*  
70: gb\_est40: \*  
71: gb\_est41: \*  
72: gb\_est42: \*  
73: gb\_est43: \*  
74: gb\_est44: \*  
75: em\_est31: \*  
76: em\_est32: \*  
77: em\_est33: \*  
78: em\_est34: \*  
79: gb\_est45: \*  
80: gb\_est46: \*  
81: gb\_est47: \*  
82: gb\_gss1: \*  
83: gb\_gss2: \*  
84: gb\_gss3: \*  
85: gb\_gss4: \*  
86: em\_gss1: \*  
87: em\_gss2: \*  
88: em\_gss3: \*  
89: em\_gss4: \*  
90: gb\_gss5: \*  
91: gb\_gss6: \*  
92: gb\_gss7: \*  
93: gb\_gss8: \*  
94: gb\_gss9: \*  
95: em\_gss5: \*  
96: em\_gss6: \*  
97: em\_gss7: \*  
98: em\_gss8: \*  
99: em\_gss9: \*  
100: em\_gss10: \*  
101: em\_gss11: \*  
102: gb\_gss10: \*  
103: gb\_gss11: \*  
104: em\_gss12: \*  
105: gb\_gss12: \*  
106: gb\_gss13: \*  
107: gb\_gss14: \*  
108: gb\_gss15: \*  
109: gb\_gss16: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result

Query

SUMMARIES

No.	Score	Match	Length	DB	ID	Description
1	12.6	74.1	75	36	AA607795	AA607795 vo49f02.r
2	12.6	74.1	98	35	D63753	D63753 D63753 Huma
3	12.2	71.8	79	43	A1191144	A1191144 gel8el2.x
4	11.2	65.9	36	21	T61825	T61825 yb92a04.s1
5	11.2	65.9	89	42	A1159672	A1159672 ut03a03.x
6	11.2	65.9	96	46	A1139657	A1139657 lc91g03.x
7	11	64.7	67	41	A1033279	A1033279 ow74b09.s
8	11	64.7	89	64	AM059703	AM059703 AHUth.bss
9	10.6	62.4	46	41	A1035942	A1035942 ub49f12.r
10	10.6	62.4	49	37	AA717442	AA717442 vu19d06.r
11	10.6	62.4	67	39	AA906835	AA906835 OK3f07.s
12	10.6	62.4	67	42	A1094466	A1094466 ou87f09.s
13	10.6	62.4	85	61	A1827573	A1827573 wf10e08.x
14	10.6	62.4	100	39	AA862553	AA862553 oh44b08.s
15	10.6	62.4	100	43	A1188636	A1188636 gdl5ell.x
16	10.6	62.4	100	64	AM065340	AM065340 687009h07
17	10.6	62.4	100	83	HOMAN	L48753 Homo sapien
18	10.2	60.0	49	37	AA677795	AA677795 z124d03.s
19	10.2	60.0	49	38	AA781646	AA781646 a150a01.s
20	10.2	60.0	58	40	AA958617	AA958617 ua15f12.r
21	10.2	60.0	61	23	H46784	H46784 y014b07.r1
22	10.2	60.0	67	44	A1273788	A1273788 qu08c02.x
23	10.2	60.0	70	29	AA177604	AA177604 mt25g10.r
24	10.2	60.0	71	61	A1869405	A1869405 tw40c01.x
25	10.2	60.0	72	37	AA675648	AA675648 vrf6f11.s
26	10.2	60.0	74	38	AA785979	AA785979 l8b10a1.f
27	10.2	60.0	76	40	AA916262	AA916262 ol63a07.s
28	10.2	60.0	82	36	AA656248	AA656248 vv12d08.r
29	10.2	60.0	82	40	AA988047	AA988047 os12g04.s
30	10.2	60.0	82	61	A1874724	A1874724 ul22e03.x
31	10.2	60.0	85	35	AA536922	AA536922 vj73h10.r
32	10.2	60.0	87	45	AJ237210	AJ237210 AJ237210
33	10.2	60.0	91	27	AA032459	AA032459 m13a11.r
34	10.2	60.0	91	30	AA219896	AA219896 mx96d11.r
35	10.2	60.0	91	74	AA214748	AA214748 uo99d06.y
36	10.2	60.0	95	37	AA701524	AA701524 z167a09.s
37	10.2	60.0	97	24	H81252	H81252 yu73g01.r1
38	10.2	60.0	97	26	W36629	W36629 mb67c04.r1
39	10.2	60.0	98	29	AA161637	AA161637 MBARG0A1
40	10	58.8	66	22	R73950	R73950 yj98h04.r1
41	10	58.8	67	62	A1877597	A1877597 fc49c10.y
42	10	58.8	84	23	H55705	H55705 CHR220644.C
43	10	58.8	85	39	AA822902	AA822902 vp18h03.r
44	10	58.8	88	35	AA574321	AA574321 nf45f03.s
45	10	58.8	91	34	AA523075	AA523075 n166c03.s

## ALIGNMENTS

RESULT 1  
AA607795 LOCUS 75 bp mRNA EST 30-SEP-1997  
DEFINITION vo49f02.r1 Barstead mouse irradiated colon MRLB7 Mus musculus CDNA  
ACCESSION AA607795  
VERSION AA607795.1 GI:2455230  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS Matra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The Washu-HHMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Apr 14, 1993 this sequence version replaced gi:692809.

Contact: Maria M/Mouse EST Project  
Washu-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:584819  
Putative full length read  
vector to vector length is 168  
Seq primer: -28m13 rev2 Et from Amersham  
High quality sequence stop: 67.  
Location/Qualifiers  
1. 75  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1053243"  
/clone\_1lb="Barstead mouse irradiated colon MRLB7"  
/dev\_stage="8 weeks"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: EcoRI; Site\_2: NotI; Tissue obtained  
from 8 week old mouse. Colon was harvested 72 hours after  
irradiation with 1400 Gys. 1st strand cDNA was primed  
with a Not I - oligo(dT) primer  
[5'-GTTTACGATCTGACGTGACGACGCGCCCTTTTCTTTTCTTTTCTTTT  
T3'] ; double-stranded cDNA was ligated to Eco RI  
adaptors [AATTCGATCTTG], digested with Not I and cloned  
into the Not I and Eco RI sites of the modified pT73  
vector. Library constructed by Bob Barstead."

BASE COUNT 21 a 19 c 20 g 15 t

Query Match 74.1%; Score 12.6; DB 36; Length 75;  
Best Local Similarity 70.6%; Pred. No. 6.5e+02;  
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaycncarcctatgt 17  
||| ||| ||| ||| |||  
Db 18 GAGGCCCAACCCATGCT 34

RESULT 2  
D63753/c LOCUS 98 bp mRNA EST 18-NOV-1997  
DEFINITION D63753 Human chromosome 11q23 mRNA (M.Katoh) Homo sapiens cDNA  
ACCESSION D63753  
VERSION D63753.1 GI:2326278  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
AUTHORS Katoh,M., Nakagawa,Y., Yawata,T., Kumano,S., Kobayashi,E., Kurimasa,A., Kugoh,H. and Oshimura,M.  
Cosmids and transcribed sequences from chromosome 11q23  
Jpn. J. Hum. Genet. 40 (4), 307-317 (1995)  
97004453  
On Sep 12, 1996 this sequence version replaced gi:1394933.  
Contact: Motonobu Katoh  
Faculty of Medicine, Department of Molecular and Cell Genetics  
Tottori University  
86 Nishimachi, Yonago, Tottori 683, Japan  
Email: mkato@grape.med.tottori-u.ac.jp.

FEATURES  
source 1. 98  
/organism="Homo sapiens"

/db.xref="taxon:9606"  
/map="11q23"  
/clone="PEX4A-3"  
/clone\_lib="Human chromosome 11q23 mRNA (M. Katoh)"

BASE COUNT 35 a 21 c 21 g 21 t

ORIGIN

Query Match 74.1%; Score 12.6; DB 35; Length 98;  
Best Local Similarity 70.6%; Pred. No. 7e+02;  
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 gaggcncarccontaygt 17  
||:|||||:|||||

Db 38 GACGCTCATCCATGCT 22

RESULT 3  
A1191144 79 bp mRNA EST 29-OCT-1998  
LOCUS A1191144  
DEFINITION gclbe12.x1 Soares\_fetal\_lung\_NbHL19W Homo sapiens cDNA clone  
IMAGE:1739374 3' similar to TR:000762 000762 UBQUITIN-CONJUGATING  
ENZYME E2; mRNA sequence.  
ACCESSION A1191144  
VERSION A1191144.1 GI:3742353  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 79)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL On Jan 17, 1998 this sequence version replaced gi:1900121.  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Trace considered overall poor quality  
Insert Length: 836 Std Error: 0.00  
Seq primer: -400P from G1bco  
High quality sequence stop: 1.  
Location/Qualifiers  
1..79  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1739374"  
/clone\_lib="Soares\_fetal\_lung\_NbHL19W"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: lung; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer-  
15'-GTGTACCAATCTGAGTGAGGAGCGCCGCAATTTTCTTTT-3',  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaldo. This library was constructed  
from the same fetus as the fetal heart library, Soares  
fetal heart NbHL19W."

BASE COUNT 12 a 23 c 22 g 22 t

ORIGIN

Query Match 71.8%; Score 12.2; DB 43; Length 79;  
Best Local Similarity 64.7%; Pred. No. 1.2e+03;  
Matches 11; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 gaggcncarccontaygt 17  
||:|||||:|||||

Db 41 GATGCCAGGCTTATGT 57

RESULT 4  
T61825 36 bp mRNA EST 14-FEB-1995  
LOCUS T61825/c  
DEFINITION yb92a04.s1 StrataGene liver (#937224) Homo sapiens cDNA clone  
IMAGE:78606 3' similar to gb:U11566 60S RIBOSOMAL PROTEIN L18  
(HUMAN); mRNA sequence.  
ACCESSION T61825  
VERSION T61825.1 GI:665068  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 36)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,  
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Roehlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevisakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert size: 693  
High quality sequence starts: 1 High quality sequence stops: 1  
Source: IMAGE Consortium, LNL This clone is available royalty-free  
through LNL; contact the IMAGE Consortium (info@image.llnl.gov)  
for further information. Trace considered overall poor quality  
Insert Length: 693 Std Error: 0.00  
Seq primer: -21m13  
High quality sequence stop: 1.  
Location/Qualifiers  
1..36  
/organism="Homo sapiens"  
/db\_xref="GDB:498351"  
/db\_xref="taxon:9606"  
/clone="IMAGE:78606"  
/clone\_lib="StrataGene liver (#937224)"  
/sex="male"  
/dev\_stage="49 years old"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: liver; Vector: pluscript SK; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. Hepatectomy from normal male caucasian. Average  
insert size: 1.1 kb; Uni-ZAP XR Vector; -5' adaptor  
sequence: 5' GAATTCGCGCAGAG 3' -3' adaptor sequence: 5'  
CTCAGATTCTTTTCTTTT 3'"

BASE COUNT 3 a 9 c 15 g 7 t 2 others

ORIGIN

Query Match 65.9%; Score 11.2; DB 21; Length 36;  
Best Local Similarity 62.5%; Pred. No. 3.9e+03;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 agygcncarccontaygt 17  
||:|||||:|||||

Db 31 ACGCCAAACCCATGCT 16

RESULT 5

A1159672/c  
 LOCUS A1159672 89 bp mRNA EST 02-OCT-1998  
 DEFINITION ufo3a05.x1 Sugano mouse liver mlia Mus musculus cDNA clone  
 IMAGE:149504 3' similar to gb:X04325.cdsl GAP JUNCTION BETA-1  
 PROTEIN (HUMAN); gb:M63802 Mus musculus connexin 32 (MOUSE);, mRNA  
 sequence.  
 ACCESSION A1159672 GI:3692854  
 VERSION A1159672  
 KEYWORDS house mouse.  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 89)  
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
 TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT On Jan 9, 1998 this sequence version replaced gl:936225.  
 Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:937108  
 Trace considered overall poor quality  
 Seg primer: custom primer used  
 High quality sequence stop: 1.

FEATURES  
 source location/Qualifiers  
 1..89  
 /organism="Mus musculus"  
 /strain="C57BL"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:149504"  
 /clone\_lib="Sugano mouse liver mlia"  
 /sex="female"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Organ: liver; Vector: pME18S-FL3; Site:1: DraIII  
 (CACTGCTG); Site:2: DraIII (CACCATGCTG); 1st strand cDNA  
 was primed with an oligo(dT) primer  
 [ATGTGGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was  
 ligated to a DraIII adaptor [TGTGGCCTACTGc], digested  
 and cloned into distinct DraIII sites of the pME18S-FL3  
 vector (5' site CACTGCTG, 3' site CACCATGCTG). XhoI should  
 be used to isolate the cDNA insert. Size selection was  
 performed to exclude fragments <1.5kb. Library  
 constructed by Dr. Sumio Sugano (University of Tokyo  
 Institute of Medical Science). Custom primers for  
 sequencing: 5' end primer CTTCGCTCTTAAGAAGCTGG and 3' end  
 primer CGACTGCGAGCTCGACGACA."

BASE COUNT 16 a 24 c 26 g 23 t  
 ORIGIN

Query Match 65.9%; Score 11.2; DB 42; Length 89;  
 Best Local Similarity 62.5%; Pred. No. 5.1e+03;  
 Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 Oy 1 gaggcncarcntayg 16  
 ||:|| ||:|| ||:||  
 Db 19 GAGCTCACACCTCATG 4

RESULT 6  
 A1439657

LOCUS A1439657 96 bp mRNA EST 28-MAR-1999  
 DEFINITION tc91g03.x1 NCI-CGAP\_C1L1 Homo sapiens cDNA clone IMAGE:2073556 3'  
 similar to gb:X57138.nml HISTONE H2B.2 (HUMAN);, mRNA sequence.  
 ACCESSION A1439657  
 VERSION A1439657  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 96)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jun 5, 1998 this sequence version replaced gl:3189048.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,  
 M.D., Louis M. Staudt, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 www-bio.lnl.gov/db/rp/image/image.html

Trace considered overall poor quality  
 Insert Length: 2241 Std Error: 0.00  
 Seq primer: -40up from G1bco  
 High quality sequence stop: 1.

FEATURES  
 source location/Qualifiers  
 1..96  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2073556"  
 /clone\_lib="NCI-CGAP\_C1L1"  
 /tissue\_type="B-cell, chronic Lymphocytic leukemia"  
 /lab\_host="DH10B"  
 /note="Vector: p1773D-Pac (Pharmacia) with a modified  
 polylinker; Site:1: Not I; Site:2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5'  
 TGTACATCTGAGTGGAGCGGCCGCGATGCTTTTCTTTTCTTTT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified p1773 vector.  
 Library is normalized, and was constructed by Bento  
 Soares and M.Fatima Bonaldo."

BASE COUNT 18 a 29 c 29 g 20 t  
 ORIGIN  
 Query Match 65.9%; Score 11.2; DB 46; Length 96;  
 Best Local Similarity 62.5%; Pred. No. 5.2e+03;  
 Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 Oy 1 gaggcncarcntayg 16  
 ||:|| ||:|| ||:|| ||:||  
 Db 79 GAGCGACAGCCGCTGTG 94

RESULT 7  
 A1033279  
 LOCUS A1033279 67 bp mRNA EST 28-AUG-1998  
 DEFINITION ow74b09.s1 Soares\_fetal\_liver\_spleen\_influs.S1 Homo sapiens CDNA  
 clone IMAGE:1652537 3', mRNA sequence.  
 ACCESSION A1033279  
 VERSION A1033279.1 GI:3254232  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;



REFERENCE	Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS	1 (bases 1 to 67) NCI-CCAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	On Jan 19, 1998 this sequence version replaced gi:2282177. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov This clone is available royalty-free through LNL ; contact the IMAGE Consortium ( <a href="mailto:info@image.llnl.gov">info@image.llnl.gov</a> ) for further information. Insert length: 553 Std Error: 0.00 Seq primer: -40ml3 fwd. ET from Amersham High quality sequence stop: 54. Location/Qualifiers
FEATURES	
source	1. .67 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1652537" /clone_id="Soares_fetal_liver_spleen_mfLS_S1" /sex="male" /dev_stage="20 week-post conception fetus" /lab_host="DH10B (ampicillin resistant)" /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; This is a subtracted version of the original Soares fetal liver spleen mFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACGGAAGATTTATTTAAGATCTTTTCTTTTCTTTTCTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT	20 a 15 c 12 g 20 t
ORIGIN	
Query Match	64.7%; Score 11; DB 41; Length 67;
Best Local Similarity	64.7%; Pred. No. 6.2e+03;
Matches 11: Conservative	2; Mismatches 4; Indels 0; Gaps 0;
0y 1 gaygcncarcncntaygt 17   :    :            :    :	
Db 31 GATGCTCGACCCCTAGT 47	
RESULT 8	
AM059703	
LOCUS	89 bp mRNA EST 16-FEB-2000
DEFINITION	AH07H.bsst.dnc15.ba.A020h09 DMC15 Homo sapiens cDNA, mRNA sequence.
ACCESSION	AM059703
VERSION	AM059703.1 GI:6652025
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS	1 (bases 1 to 89) Brenner,S., Williams,S.R., Vermass,E.H., Storch,T., Moon,K., McCollum,C., Mao,J.I., Kirchner,J.J., Eletre,S., Dubridge,R.B., Burcham,T., and Albrecht,G.
TITLE	In vitro cloning of complex mixtures of DNA on microbeads: Physical separation of differentially expressed cDNAs
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (4), 1665-1670 (2000)
COMMENT	On Jun 5, 1998 this sequence version replaced gi:318342. Contact: Burcham TS LYNX Therapeutics, Inc. 25861 Industrial Blvd., Hayward, CA 94545, USA Tel: 510 670 9338 Fax: 510 670 9302 Email: <a href="mailto:tlmb@lynxgen.com">tlmb@lynxgen.com</a>

Sequence obtained from LMXN Therapeutics Megasort technology.  
Collected from the down-regulated gate.  
High quality sequence stop: 89.  
Location/Qualifiers  
1. . 89

FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="DNC15"  
/cell\_type="monocytic leukemia"  
/note="Vector: PCR2.1; Cloning of PCR products from micro-beads carrying 3' end of down-regulated cDNA. THP-1 cells non-induced (treated with DMSO only)."

BASE COUNT  
21 a 21 c 24 g 23 t

ORIGIN

Query Match 64.7%; Score 11; DB 64; Length 89;  
Best Local Similarity 64.7%; Pred. No. 6.8e+03;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Oy 1 gagcncarcclataygt 17  
||:| | | | | | | |  
Db 41 GACGCACATACATACGT 57

RESULT 9  
AI035942/c

LOCUS  
DEFINITION  
ub99f12.r1 Soares\_mammary\_gland\_NbMWG Mus musculus cDNA clone  
IMAGE:1381103 5' similar to TR:Q99969 Q99969 TAZAROTENE-INDUCED  
GENE 2. ; mRNA sequence.  
AI035942  
AI035942.1 GI:3259633  
EST.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 46)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
On Sep 12, 1996 this sequence version replaced gi:1400925.  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MG1:903571

REFERENCE  
AUTHORS

TITLE  
JOURNAL  
COMMENT

FEATURES  
source  
1. 46  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1381103"  
/clone\_lib="Soares\_mammary\_gland\_NbMWG"  
/sex="male"  
/tissue\_type="mammary gland"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"

```

/note="Organ: mammary gland; Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
TGTTCACATCTGAGTGGAGCGCCCGCATGCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."

BASE COUNT      8 a      12 c      15 g      11 t
ORIGIN

Query Match      62.4%; Score 10.6; DB 41; Length 46;
Best Local Similarity 66.7%; Pred. No. 9.9e+03;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy      3 ygcncarcntaygt 17
      :|||:|||:|||
      Db      38 TGCCCGACCATAGCT 24

RESULT 10
LOCUS      AA1717442      49 bp      mRNA      EST      29-DEC-1997
DEFINITION vuj9d06.r1 Barsstead mouse myotubes MPLRB5 Mus musculus cDNA clone
IMAGE:1181099 5' similar to TR:000308 000308 WMP2.;, mRNA
sequence.
AA1717442      AA1717442.1 GI:2729716
VERSION      AA1717442
KEYWORDS      house mouse.
SOURCE      Mus musculus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
      Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 49)
AUTHORS      Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
      Gesel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
      Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
      Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
      Waterston,R.
      The WashU-HHMI Mouse EST Project
      Unpublished (1996)
      On May 5, 1995 this sequence version replaced gi:798372.
      Contact: Marra M/Mouse EST Project
      WashU-HHMI Mouse EST Project
      Washington University School of Medicinep
      4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
      Tel: 314 286 1800
      Fax: 314 286 1810
      Email: mouseest@wustl.wustl.edu
      This clone is available royalty-free through LNL; contact the
      IMAGE Consortium (info@image.llnl.gov) for further information.
      MGI:638947
      Trace considered overall poor quality
      Possible reversed clone; similarity on wrong strand
      Seq primer: -28m13 rev2 ET from Amersham
      High quality sequence stop: 1.
      Location/Qualifiers
      1..49
      /organism="Mus musculus"
      /strain="C3H"
      /db_xref="taxon:10090"
      /clone IMAGE:1181099"
      /clone_1lb="Barsstead mouse myotubes MPLRB5"
      /cell_line="C2C12"
      /lab_host="DH10B"
      /note="Vector: pT73D-Pac (Pharmacia) with a modified
      polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
      was primed with a Not I - oligo(dT) primer [5'
      TGTTCACATCTGAGTGGAGCGCCCGCATGCTTTTCTTTTCTTTTCTTTT

```

```

3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATCGATCTCTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barsstead. The C2C12 cell line
(available from ATCC, catalog # CRL-1772) differentiates
rapidly, forming contractile myotubes and producing
characteristic muscle proteins."

BASE COUNT      7 a      12 c      12 g      18 t
ORIGIN

Query Match      62.4%; Score 10.6; DB 37; Length 49;
Best Local Similarity 58.8%; Pred. No. 1e+04;
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy      1 gaycncarcntaygt 17
      :|||:|||:|||
      Db      19 GGTCCTGACCTATGT 35

RESULT 11
LOCUS      AA906835/C      67 bp      mRNA      EST      09-JUN-1998
DEFINITION OK73F07.s1 NCI-CGAP_GC4 Homo sapiens cDNA clone IMAGE:1519621 3'
IMAGE:1519621 5' similar to SW:ARSE_HUMAN P51690 ARYL5ULFATASE E PRECURSOR.;, mRNA
sequence.
AA906835      AA906835.1 GI:3042079
VERSION      AA906835
KEYWORDS      human.
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
      Eutheria; Primates; Catarrhini; Homindae; Homo.
REFERENCE      1 (bases 1 to 67)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
      Tumor Gene Index
      Unpublished (1997)
      On Jan 14, 1998 this sequence version replaced gi:1798604.
      Contact: Robert Strausberg, Ph.D.
      Tel: (301) 496-1550
      Email: Robert.Strausberg@nih.gov
      Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
      Emmert-Buck, M.D., Ph.D.
      cDNA Library Preparation: M. Bento Soares, Ph.D.
      DNA Sequencing by: Washington University Genome Sequencing Center
      Clone distribution: NCI-CGAP clone distribution information can be
      found through the I.M.A.G.E. Consortium/LLNL at:
      www-bio.llnl.gov/dbp/image/image.html

Trace considered overall poor quality
Insert Length: 824 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.
Location/Qualifiers
1..67
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone IMAGE:1519621"
/clone_1lb="NCI-CGAP_GC4"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from 3 pooled
germ cell tumors, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."

BASE COUNT      19 a      13 c      20 g      15 t
ORIGIN

```

QY	1	gaygcncarcnctaygt	17
DB	38	GAGCGACGATCCTTAC	24
Query Match	62.4%	Score 10.6;	DB 39; Length 67;
Best Local Similarity	66.7%	Pred. No. 1.le+04;	
Matches	10;	Conservative	2; Mismatches 3; Indels 0; Gaps 0;
RESULT	12		
LOCUS	A1094466/c		
DEFINITION	oab7f09.s1 Soares NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone IMAGE1634825.3 similar to contains MER22.t2 MER22 repetitive element.; mRNA sequence.		
VERSION	A1094466		
KEYWORDS	A1094466.1 GI:3433442		
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 67)		
JOURNAL	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.		
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
Unpublished (1997)			
On Jan 19, 1998 this sequence version replaced gi:2284745.			
Contact: Robert Strausberg, Ph.D.			
Tel: (301) 496-1550			
Email: Robert.Strausberg@nih.gov			
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (infoimage.lnl.gov) for further information.			
Insert Length: 431 Std Error: 0.00			
Seq primer: -40ml3 fwd. EF from Amersham			
High quality sequence stop: 53.			
location/Qualifiers			
1..67			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/clone="IMAGE:1634825"			
/clone_lid="Soares_NSF_F8_9W_OT_PA_P_S1"			
/lab_host="DH10B"			
/note="Organ: pooled; Vector: pT773D-Pac (Pharmacia) with a modified polylinker; Site:1: Not I; Site:2: Eco RI; Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles was made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and cloneids: Soares NBHSF pool 1: 309384-310919, 333208-325895 Soares Nb2HP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares Nb2HR8-9W pool 1: 758280-760583, 772104-774407 Soares NBHP pool 1: 304776-3206311, 320136-322823, 326280-326653 Soares NBHOT pool 1: 723720-726407, 739080-740999 Subtraction by Benito Soares and M. Fatima Bernaldo."			
BASE COUNT	23 a	13 c	13 t
ORIGIN			
Query Match	62.4%	Score 10.6;	DB 42; Length 67;
Best Local Similarity	58.8%	Pred. No. 1.le+04;	
Matches	10;	Conservative	3; Mismatches 4; Indels 0; Gaps 0;
QY	1	gaygcncarcnctaygt	17
DB	57	GATGCCAGCCCAACGT	41

RESULT	13	
LOCUS	AI827573	85 bp mRNA EST 18-DEC-1999
DEFINITION	w110e08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2350214 3' similar to TR:000605 000605 PUTATIVE RAB5-INTERACTING PROTEIN (CLONE L1-57) ;, mRNA sequence.	
ACCESSION	AI827573	
VERSION	AI827573.1	GI:5448244
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 85)	
AUTHORS	NCI-CCAG <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
JOURNAL	Unpublished (1997)	
COMMENT	On Jun 5, 1998 this sequence version replaced gi:3188016. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert_Strausberg@nih.gov This clone is available royally-free through LNL ; contact the IMAGE Consortium ( <a href="mailto:info@image.flnl.gov">info@image.flnl.gov</a> ) for further information. Trace considered overall poor quality Insert Length: 1130 Std Error: 0.00 Seq primer: -400P from Glibco High quality sequence stop: 1. Location/Qualifiers	
FEATURES	1..85	
source	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2350214" /clone_lib="Soares_NFL_T_GBC_S1" /lab_host="DH10B" /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBH139W, testis NHR, and B-cell NCL1CGAP-GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."	
BASE COUNT	16 a 22 c 20 g 27 t	
ORIGIN		
Query Match	62.4%;	Score 10.6; DB 61; Length 85;
Best Local Similarity	66.7%;	Pred. No. 1.2e+04;
Matches 10; Conservative	2; Mismatches 3; Indels 0; Gaps 0;	
OY	3 ygcncarcntaygt 17	
Db	36 TGCATCATCCATATGTT 50	
RESULT	14	
LOCUS	AA862553	100 bp mRNA EST 24-AUG-1998
DEFINITION	OH4db08.s1 NCI-CCAG.GC4 Homo sapiens cDNA clone IMAGE:1469463 3' similar to SW:DDX4_RAT Q64060 DEAD BOX PROTEIN 4 ;, mRNA sequence.	
ACCESSION	AA862553	
VERSION	AA862553.1	GI:2955032
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	

REFERENCE 1 (bases 1 to 100)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1402122.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbcrp/image/image.html](http://www.bio.llnl.gov/dbcrp/image/image.html)

Trace considered overall poor quality  
Insert Length: 1202 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers

FEATURES  
source  
1..100  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1469463"  
/clone.lib="NCI-CGAP\_GC4"  
/tissue.type="pooled germ cell tumors"  
/lab.host="DH10B"  
/note="Vector: p773D-Pac (Pharmacia) with a modified  
polylinker; 1st strand cDNA was prepared from 3 pooled  
germ cell tumors, and was then primed with a Not I -  
oligo(dT) primer. Double-stranded cDNA was ligated to Eco  
RI adaptors (Pharmacia), digested with Not I and cloned  
into the Not I and Eco RI sites of the modified p773  
vector. Library is normalized. Library was constructed by  
Bento Soares and M. Fatima Bonaldo."

BASE COUNT 30 a 18 c 19 g 33 t  
ORIGIN

Query Match 62.4%; Score 10.6; DB 39; Length 100;  
Best Local Similarity 58.8%; Pred. No. 1.2e+04;  
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 gaygcncarcctaygt 17  
|:|:| |:| |:|  
Db 58 GATGCTCAACATGATGT 42

RESULT 15  
LOCUS A1188636 100 bp mRNA EST 28-OCT-1998  
DEFINITION qd15ell.x1 Soares-placenta\_8to9weeks\_2NBHP8co9w Homo sapiens CDNA  
clone IMAGE:1723820 3' similar to TR:Q99519 Q99519 SIALIDASE  
PRECURSOR.; mRNA sequence.  
ACCESSION A1188636  
VERSION A1188636.1 GI:3739845  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 100)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LLNL; contact the

IMAGE Consortium ([info@image.llnl.gov](http://info@image.llnl.gov)) for further information.  
Trace considered overall poor quality  
Insert Length: 859 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 1.  
Location/Qualifiers

FEATURES  
source  
1..100  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1723820"  
/clone.lib="Soares.placenta\_8to9weeks\_2NBHP8co9w"  
/dev.stage="two placenta: one from 8 weeks and another  
from 9 weeks post conception"  
/lab.host="DH10B (ampicillin resistant)"  
/note="Organ: placenta; Vector: p773D (Pharmacia) with a  
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTCACATCTGAGGTGGAGCGCGCGGATTTTGTTTTGT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified p773 vector  
(Pharmacia). Library constructed by Bento Soares and  
M. Fatima Bonaldo."

BASE COUNT 18 a 28 c 28 g 26 t  
ORIGIN

Query Match 62.4%; Score 10.6; DB 43; Length 100;  
Best Local Similarity 58.8%; Pred. No. 1.2e+04;  
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 gaygcncarcctaygt 17  
|:|:| |:| |:|  
Db 88 GCCGCCACGCTGTACGT 72

Search completed: June 4, 2000, 13:52:56  
Job time: 20920 sec



**THIS PAGE BLANK (USPTO)**



```
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,318
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/446,872
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 100:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 68 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic DNA"
US-08-469-318-100
```

Query Match 69.4%; Score 11.8; DB 5; Length 68;  
Best Local Similarity 66.7%; Pred. No. 1.5e+02;  
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 ygcncarcntaygt 17  
:|||:|||||:  
Db 20 TGGCGAGCCCTACGT 6

```
RESULT 3
PCT-US95-01185-99
; Sequence 99, Application PC/TUS9501185
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Multivariant IL-3 Hematopoiesis Fusion
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 196
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01185
; FILING DATE: 02-FEB-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/192325
; FILING DATE: 14-FEB-1994
; INFORMATION FOR SEQ ID NO: 99:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 68 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic DNA"
PCT-US95-01185-99
```

Query Match 69.4%; Score 11.8; DB 6; Length 68;  
Best Local Similarity 66.7%; Pred. No. 1.5e+02;  
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 ygcncarcntaygt 17  
:|||:|||||:  
Db 53 TGGCGAGCCCTACGT 67

RESULT 4

```
PCT-US95-01185-100/C
; Sequence 100, Application PC/TUS9501185
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Multivariant IL-3 Hematopoiesis Fusion
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 196
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01185
; FILING DATE: 02-FEB-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/192325
; FILING DATE: 14-FEB-1994
; INFORMATION FOR SEQ ID NO: 100:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 68 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic DNA"
PCT-US95-01185-100
```

Query Match 69.4%; Score 11.8; DB 6; Length 68;  
Best Local Similarity 66.7%; Pred. No. 1.5e+02;  
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 ygcncarcntaygt 17  
:|||:|||||:  
Db 20 TGGCGAGCCCTACGT 6

```
RESULT 5
US-08-852-807-28/C
; Sequence 28, Application US/08852807
; Patent No. 5861298
; GENERAL INFORMATION:
; APPLICANT: Dehouck, Christine
; APPLICANT: Drake, Fred
; APPLICANT: Gowen, Maxine
; APPLICANT: Rood, Julie
; APPLICANT: Hastings, Gregg
; APPLICANT: Adams, Mark
; APPLICANT: Fraser, Claire
; APPLICANT: Lee, No. 5861298man
; APPLICANT: Kirkness, Ewen
; APPLICANT: Blake, Judith
; APPLICANT: Fitzgerald, Lisa
; TITLE OF INVENTION: CATHEPSIN K GENE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Smithkline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-2799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,807
; FILING DATE: 07-MAY-1997
; CLASSIFICATION: 435
```



```

1      PRIOR APPLICATION DATA:
2      APPLICATION NUMBER: 60/019,942
3      FILING DATE: 14-JUNE-1996
4      PRIOR APPLICATION DATA:
5      APPLICATION NUMBER: 60/020,273
6      FILING DATE: 17-JUNE-1996
7      PRIOR APPLICATION DATA:
8      APPLICATION NUMBER: 60/026,273
9      FILING DATE: 26-AUG-1996
10     ATTORNEY/AGENT INFORMATION:
11     NAME: Han, William T
12     REGISTRATION NUMBER: 34,344
13     REFERENCE/DOCKET NUMBER: ATG50006-2
14     TELECOMMUNICATION INFORMATION:
15     TELEPHONE: 610-270-5219
16     TELEFAX: 610-270-5090
17     TELEX:
18
19     INFORMATION FOR SEQ ID NO: 28:
20     SEQUENCE CHARACTERISTICS:
21     LENGTH: 21 base pairs
22     TYPE: nucleic acid
23     STRANDEDNESS: single
24     TOPOLOGY: linear
25     MOLECULE TYPE: Genomic DNA
26     HYPOTHEICAL: NO
27     ANTI-SENSE: NO
28     FRAGMENT TYPE:
29     ORIGINAL SOURCE:
30     US-08-852-807-28

```

	64.7%;	Score 11;	DB 3;	Length 21;
Query Match				
Best Local Similarity	64.7%;	Pred. No.	4.2e+02;	
Matches 11;	Conservative 2;	Mismatches 4;	Indels 0;	Gaps 0;
OY	1 gavgcncarcntatgct	17		
	:			
db	18 GATGCGCTACCCATTATGT	2		

RESULT 6  
 US-08-852-807-29  
 Sequence 29, Application US/08852807  
 Patent No. 5861238  
 GENERAL INFORMATION:  
 APPLICANT: Debuock, Christine  
 APPLICANT: Drake, Fred  
 APPLICANT: Gomen, Maxine  
 APPLICANT: Rood, Julie  
 APPLICANT: Hastings, Gregg  
 APPLICANT: Adams, Mark  
 APPLICANT: Fraser, Claire  
 APPLICANT: Lee, No. 5861298man  
 APPLICANT: Kirkness, Ewen  
 APPLICANT: Blake, Judith  
 APPLICANT: Fitzgerald, Lisa  
 TITLE OF INVENTION: CATHEPSIN K GENE  
 NUMBER OF SEQUENCES: 32  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Smithkline Beecham Corporation  
 STREET: 709 Swedeland Road  
 CITY: King of Prussia  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19406-2799  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSEO version 1.5  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/852,807  
 FILING DATE: 07-MAY-1997

```

1 CLASSIFICATION: 435
2 PRIOR APPLICATION DATA:
3 APPLICATION NUMBER: 60/019, 942
4 FILING DATE: 14-JUNE-1996
5 PRIOR APPLICATION DATA:
6 APPLICATION NUMBER: 60/020, 273
7 FILING DATE: 17-JUNE-1996
8 PRIOR APPLICATION DATA:
9 APPLICATION NUMBER: 60/026, 273
10 FILING DATE: 26-AUG-1996
11 ATTORNEY/AGENT INFORMATION:
12 NAME: Han, William T
13 REGISTRATION NUMBER: 34,344
14 REFERENCE/DOCKET NUMBER: ATG50006-2
15 TELECOMMUNICATION INFORMATION:
16 TELEPHONE: 610-270-5219
17 TELEFAX: 610-270-5090
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
100
```

Query Match	64.7%	Score 11;	DB 3;	Length 21;
Best Local Similarity	64.7%	Pred. No. 4.2e+02;		
Matches 11; Conservative	2;	Mismatches 4;	Indels 0;	Gaps 0;
Qy	1	gagcncarcntatgct	17	
	:			
Db	4	gATGGCTTACCCATATGT	20	

RESULT 7  
 US-08-850-049-18  
 Sequence 18 Application US/08850049  
 Patent No. 5965726  
 GENERAL INFORMATION:  
 APPLICANT:  
 APPLICANT:  
 APPLICANT:  
 TITLE OF INVENTION: METHOD OF ELIMINATING  
 TITLE OF INVENTION: INHIBITORY/INSTABILITY REGIONS OF RNA  
 NUMBER OF SEQUENCES: 130  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: MORGAN & FINNEGAN  
 STREET: 345 PARK AVENUE  
 CITY: NEW YORK  
 STATE: NEW YORK  
 COUNTRY: USA  
 ZIP: 10154  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: FLOPPY DISK  
 COMPUTER: IBM PC COMPATIBLE  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: WORDPERECT 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/850,049  
 FILING DATE: 02-MAY-1997  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/050,478  
 FILING DATE: 26-OCT-1994  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/02908  
FILING DATE: 29-MAR-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/858,747  
FILING DATE: 27-MAR-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MORRY, MARY J.  
REGISTRATION NUMBER: 34,398  
REFERENCE/DOCKET NUMBER: 2026-4006US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)751-6849  
TELEFAX: (212)751-6849  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 BASE PAIRS  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
US-08-850-049-18

Query Match 64.7%; Score 11; DB 4; Length 50;  
Best Local Similarity 64.7%; Pred. No. 4.4e+02;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gaygcncarcctaygt 17  
||| ||| ||| ||| |||  
Db 19 GAAGGACAGCCACTAGCT 35

RESULT 8  
US-08-050-478-18  
Sequence 18, Application US/08050478  
Patent No. 5972596  
GENERAL INFORMATION:  
APPLICANT:  
APPLICANT:  
TITLE OF INVENTION: METHOD OF ELIMINATING  
TITLE OF INVENTION: INHIBITORY/INSTABILITY REGIONS OF mRNA  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/050.478  
FILING DATE: 26-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/02908  
FILING DATE: 29-MAR-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/858,747  
FILING DATE: 27-MAR-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MORRY, MARY J.  
REGISTRATION NUMBER: 34,398  
REFERENCE/DOCKET NUMBER: 2026-4006US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)758-4800

TELEFAX: (212)751-6849  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 BASE PAIRS  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
US-08-050-478-18

Query Match 64.7%; Score 11; DB 4; Length 50;  
Best Local Similarity 64.7%; Pred. No. 4.4e+02;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gaygcncarcctaygt 17  
||| ||| ||| ||| |||  
Db 19 GAAGGACAGCCACTAGCT 35

RESULT 9  
PCT-US94-06079-41/c  
Sequence 41, Application PC/TUS9406079  
GENERAL INFORMATION:  
APPLICANT: American Cyanamid Company,  
TITLE OF INVENTION: Gene Insertion by Direct Ligation In  
TITLE OF INVENTION: Vltro  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: American Cyanamid Company  
STREET: One Cyanamid Plaza  
CITY: Wayne  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07470-8426  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/06079  
FILING DATE: 27-MAY-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Gordon, Alan M.  
REGISTRATION NUMBER: 30,637  
REFERENCE/DOCKET NUMBER: 31969-00\PCF  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-831-3244  
TELEFAX: 201-831-3305  
INFORMATION FOR SEQ ID NO: 41:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 63 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
PCT-US94-06079-41

Query Match 64.7%; Score 11; DB 6; Length 63;  
Best Local Similarity 64.7%; Pred. No. 4.5e+02;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gaygcncarcctaygt 17  
||| ||| ||| ||| |||  
Db 24 GAGGCCAGCCGACGCT 8

RESULT 10  
US-08-009-265-21/c

```
; Sequence 21, Application US/08009265
; Patent No. 5547871
; GENERAL INFORMATION:
; APPLICANT: Black Dr., Bruce C.
; APPLICANT: Summers Dr., Max D.
; TITLE OF INVENTION: Heterologous Signal Sequences For
; TITLE OF INVENTION: Secretion Of Insect Controlling Proteins
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: 1937 West Main Street, P.O. Box 60
; CITY: Stamford
; STATE: CT
; COUNTRY: USA
; ZIP: 06904-0060
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/009,265
; FILING DATE: 19930125
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gordon Mr., Alan M.
; REGISTRATION NUMBER: 30637
; REFERENCE/DOCKET NUMBER: 31868-00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 203-321-2719
; TELEFAX: 203-321-2971
; TELEX: 710-474-4059
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 75 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Drosophila melanogaster
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 13..75
; US-08-009-265-21

Query Match      64.7%; Score 11; DB 1; Length 75;
Best Local Similarity 64.7%; Pred. No. 4.6e+02;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy      1 gaycncarcncatlayt 17
      |||||:|||||
Db      36 GAGGCCAGCCCGACGT 20
```

```
RESULT 11
US-08-477-527A-94/C
; Sequence 94, Application US/08477527A
; Patent No. 5972599
; GENERAL INFORMATION:
; APPLICANT: DIANE TASSETT
; APPLICANT: NIKOS PAGRATIS
; APPLICANT: SUMEDHA JAYASENA
; APPLICANT: LARRY GOLD
; TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS
; TITLE OF INVENTION: OF CYTOKINES
; NUMBER OF SEQUENCES: 258
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson and Bratschun, L.L.C.
; STREET: 8400 East Prentice Avenue, Suite #200
```

```
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,527A
; FILING DATE: 7-JUNE-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/714,131
; FILING DATE: 10-JUNE-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/931,473
; FILING DATE: 17-AUGUST-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/964,624
; FILING DATE: 21-OCTOBER-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/117,991
; FILING DATE: 8-SEPTEMBER-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/536,428
; FILING DATE: 11-JUNE-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Diane H. McLearn
; REGISTRATION NUMBER: 33,960
; REFERENCE/DOCKET NUMBER: NEX41-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 86 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: RNA
; FEATURE:
; OTHER INFORMATION: All pyrimidines are 2'-F
; OTHER INFORMATION: modified
; US-08-477-527A-94

Query Match      64.7%; Score 11; DB 4; Length 86;
Best Local Similarity 64.7%; Pred. No. 4.6e+02;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy      1 gaycncarcncatlayt 17
      |||||:|||||
Db      65 GAAGCTCATCATCATAGT 49
```

```
RESULT 12
US-08-481-710-94/C
; Sequence 94, Application US/08481710
; Patent No. 6028186
; GENERAL INFORMATION:
; APPLICANT: DIANE TASSETT
; APPLICANT: NIKOS PAGRATIS
; APPLICANT: SUMEDHA JAYASENA
; APPLICANT: LARRY GOLD
; TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS
; TITLE OF INVENTION: OF CYTOKINES
; NUMBER OF SEQUENCES: 258
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson and Bratschun, L.L.C.
; STREET: 8400 East Prentice Avenue, Suite #200
; CITY: Denver
; STATE: Colorado
```

COUNTRY: USA  
ZIP: 80111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/481,710  
FILING DATE: 7-JUNE-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/714,131  
FILING DATE: 10-JUNE-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/931,473  
FILING DATE: 17-AUGUST-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/964,624  
FILING DATE: 21-OCTOBER-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/117,991  
FILING DATE: 8-SEPTEMBER-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/536,428  
FILING DATE: 11-JUNE-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Diane H. McLearn  
REGISTRATION NUMBER: 33,960  
REFERENCE/DOCKET NUMBER: NEX41-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 793-3333  
TELEFAX: (303) 793-3433  
INFORMATION FOR SEQ ID NO: 94:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 86 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: RNA  
FEATURE:  
OTHER INFORMATION: All pyrimidines are 2'-F  
US-08-481-710-94

Query Match 64.7%; Score 11; DB 5; Length 86;  
Best Local Similarity 64.7%; Pred. No. 4.6e+02;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 1 gaycncarcntaygt 17  
||| ||| : ||| |||  
DB 65 GAAGCTCATCATATCT 49  
RESULT 13  
PCT-US96-09537-94/C  
Sequence 94, Application PC/TUS9609537  
GENERAL INFORMATION:  
APPLICANT: NEXSTAR PHARMACEUTICALS, INC.  
APPLICANT: DIANE TASSER  
APPLICANT: NIKOS PAGRATIS  
APPLICANT: SUMEDHA JAYASENA  
APPLICANT: LARRY GOLD  
TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS OF  
NUMBER OF SEQUENCES: 258  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Swanson and Bratschun, L.L.C.  
STREET: 8400 East Prentice Avenue, Suite #200  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA

ZIP: 80111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/09537  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/477,829  
FILING DATE: 07-JUNE-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/481,710  
FILING DATE: 07-JUNE-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Barry J. Swanson  
REGISTRATION NUMBER: 33,215  
REFERENCE/DOCKET NUMBER: NEX41/PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 793-3333  
TELEFAX: (303) 793-3433  
INFORMATION FOR SEQ ID NO: 94:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 86 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: RNA  
FEATURE:  
OTHER INFORMATION: All pyrimidines are 2'-F  
PCT-US96-09537-94

Query Match 64.7%; Score 11; DB 6; Length 86;  
Best Local Similarity 64.7%; Pred. No. 4.6e+02;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 1 gaycncarcntaygt 17  
||| ||| : ||| |||  
DB 65 GAAGCTCATCATATGT 49  
RESULT 14  
US-08-009-265-39/C  
Sequence 39, Application US/08009265  
Patent No. 5547871  
GENERAL INFORMATION:  
APPLICANT: Black Dr., Bruce C.  
APPLICANT: Summers Dr., Max D.  
TITLE OF INVENTION: Heterologous Signal Sequences For  
TITLE OF INVENTION: Secretion Of Insect Controlling Proteins  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: American Cyanamid Company  
STREET: 1937 West Main Street, P.O. Box 60  
CITY: Stamford  
STATE: CT  
COUNTRY: USA  
ZIP: 06904-0060  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/009,265  
FILING DATE: 19930125  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gordon M., Alan M.

REGISTRATION NUMBER: 30637  
 REFERENCE/DOCKEN NUMBER: 31668-00  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 203-321-2719  
 TELEFAX: 203-321-2971  
 TELEX: 710-474-0539  
 INFORMATION FOR SEQ ID NO: 39:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 90 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHEetical: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Drosophila melanogaster

	Query March	64.7%	Score 11;	DB 1;	Length 90;
	Best Local Similarity	64.7%	Pred. No. 4.	6e-02;	
	Matches 11;	Conservative	2;	Mismatches 4;	Indels 0;
	Gaps	0;			
Oy	1 gaycgcncarcctataygt	17			
Db	36 GAGGCCCGACCCGACGT	20			

US-RESULT 15  
US-08-477-527A-103  
Sequence 103, Application US/08477527A  
Patent No. 5972599  
GENERAL INFORMATION:  
APPLICANT: DIANE TASSER  
APPLICANT: NIKOS PAGRATIS  
APPLICANT: SUNEDEHA JAYASENA  
APPLICANT: LARRY GOLD  
TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS  
TITLE OF INVENTION: OF CYTOKINES  
NUMBER OF SEQUENCES: 258  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Swanson and Bratschun, L.L.C.  
CITY: Denver  
STREET: 8400 East Prentice Avenue, Suite #200  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 Inch, 1.44 MB storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,527A  
FILING DATE: 7-JUNE-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/714,131  
FILING DATE: 10-JUNE-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/931,473  
FILING DATE: 17-AUGUST-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/964,624  
FILING DATE: 21-OCTOBER-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/117,991  
FILING DATE: 8-SEPTEMBER-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/536,428  
FILING DATE: 11-JUNE-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Diane H. McLearn

```

1  REGISTRATION NUMBER: 33,960
2  REFERENCE/DOCKET NUMBER: NEXA1-1
3  TELECOMMUNICATION INFORMATION:
4  TELEPHONE: (303) 793-3333
5  TELEFAX: (303) 793-3433
6  INFORMATION FOR SEQ ID NO: 103:
7  SEQUENCE CHARACTERISTICS:
8  LENGTH: 80 base pairs
9  TYPE: nucleic acid
10 STRANDEDNESS: single
11 TOPOLOGY: linear
12 MOLECULE TYPE: RNA
13 FEATURE:
14 OTHER INFORMATION: All pyrimidines are 2'-F-
15 OTHER INFORMATION: modified
16 US-08-477-527A-103

```

Query Match	63.5%	Score 10.8	DB 4	Length 80
Best Local Similarity	57.1%	Pred. NO. 6e+02		
Matches	8	Conservative	4	Mismatches 2
				Indels 0
				Gaps 0
Qy	2	aygcncarcntay	15	
		: : : : : :		
Db	16	ACGCUCACCCGUAU	29	

Search completed: June 4, 2000, 16:09:37  
Job time: 28056 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:05:23 ; Search time 1236.38 Seconds

(without alignments)  
-13.376 Million cell updates/sec

Title: US-09-164-714-3

Perfect score: 17  
Sequence: 1 tgyttingcrtcnacytg 17

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 356616

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_om:\*  
5: gb\_om:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vl:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_om:\*  
23: em\_ov:\*  
24: em\_ov:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_ro:\*  
28: em\_ro:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vl:\*  
32: gb\_htg1:\*  
33: gb\_htg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pt4:\*  
41: gb\_htg3:\*  
42: gb\_htg4:\*  
43: gb\_htg5:\*  
44: gb\_htg6:\*

45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pt5:\*  
52: gb\_htg8:\*  
53: gb\_htg9:\*  
54: gb\_htg10:\*  
55: gb\_htg11:\*  
56: gb\_htg12:\*  
57: gb\_htg13:\*  
58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	12.6	74.1	86	5	AR042759	AR042759 Sequence
2	12.6	74.1	86	5	I65650	I65650 Sequence 10
3	12.6	74.1	86	5	I67882	I67882 Sequence 10
4	12.6	74.1	86	5	I90103	I90103 Sequence 10
5	12.2	71.8	48	5	E07909	E07909 PCR primer
6	12.2	71.8	98	5	A59486	A59486 Sequence 36
7	11.2	65.9	50	5	I42237	I42237 Sequence 50
8	11.2	65.9	53	5	AR051504	AR051504 Sequence 7
9	11.2	65.9	54	5	A43617	A43617 Sequence 7
10	11.2	65.9	54	5	AR040961	AR040961 Sequence
11	11.2	65.9	57	5	AR040962	AR040962 Sequence
12	11.2	65.9	57	5	AR051866	AR051866 Sequence
13	11.2	65.9	57	5	AR051867	AR051867 Sequence
14	11.2	65.9	73	8	ATSAABR08	U19128 Arabidopsis
15	11.2	65.9	84	3	CATFES13	M16699 Feline c-fe
16	11.2	65.9	93	7	EA28SD1	Z54117 E.album 28S
17	11.2	65.9	93	8	BAU35417	U35417 Beauveria a
18	11.2	65.9	96	5	E07904	E07904 DNA fragmen
19	11.2	65.9	99	9	HUMTRA	M20911 Human tyros
20	11	64.7	21	3	AB005886	AB005886 Bos tauru
21	11	64.7	31	5	A45726	A45726 Sequence 5
22	11	64.7	80	34	DROTRO06	L00359 D.melanogas
23	10.6	62.4	20	5	A82484	A82484 Sequence 22
24	10.6	62.4	22	5	I46451	I46451 Sequence 43
25	10.6	62.4	22	5	I46452	I46452 Sequence 43
26	10.6	62.4	22	5	I46453	I46453 Sequence 43
27	10.6	62.4	24	5	A27856	A27856 Mutagenese
28	10.6	62.4	27	5	I12386	I12386 Sequence 9
29	10.6	62.4	30	5	AR004750	AR004750 Sequence
30	10.6	62.4	30	5	AR008236	AR008236 Sequence
31	10.6	62.4	30	5	I77020	I77020 Sequence 80
32	10.6	62.4	30	5	I81015	I81015 Sequence 80
33	10.6	62.4	30	5	I81111	I81111 Sequence 80
34	10.6	62.4	32	5	A83637	A83637 Sequence 4
35	10.6	62.4	36	5	I17695	I17695 Sequence 12
36	10.6	62.4	42	5	A05115	A05115 Oligonucleo
37	10.6	62.4	43	12	MMQAL1RNA	X60888 M.musculus
38	10.6	62.4	50	14	SYNP17XORC	K00453 chimeric pl
39	10.6	62.4	51	12	U92176	U92176 Mus musculin
40	10.6	62.4	56	5	E03956	E03956 DNA encodin
41	10.6	62.4	57	12	AF224185	AF224185 Mus muscu
42	10.6	62.4	59	5	E00072	E00072 Partial CDN
43	10.6	62.4	75	5	I12390	I12390 Sequence 13
44	10.6	62.4	81	16	AF018384	AF018384 Hepatitis
45	10.6	62.4	81	16	AF018385	AF018385 Hepatitis

ALIGNMENTS

[illegible]

AUTHORS	JanJic,N. and Gold,L.
TITLE	High affinity PDGF nucleic acid ligands
JOURNAL	Patent: US 5674685-A 10 07-Oct-1997;
FEATURES	Location/Qualifiers
SOURCE	1..86
BASE COUNT	22 a 21 c 25 g 18 t
ORIGIN	/organism="unknown"
Query Match	74.1%; Score 12.6; DB 5; Length 86;
Best Local Similarity	70.6%; Pred. No. 3.3e+03;
Matches 12; Conservative	2; Mismatches 3; Indels 0; Gaps 0;
OY	1 tgytngcrtcnacytg 17       :      :
Db	50 TGCTTAGCGTCACACTTG 66
RESULT 4	
LOCUS	I90103 86 bp DNA PAT 10-AUG-1998
DEFINITION	Sequence 10 from patent US 5723594.
ACCESSION	I90103
VERSION	I90103.1 GI:3410043
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 86)
TITLE	JanJic,N. and Gold,L.
JOURNAL	High affinity PDGF nucleic acid ligands
FEATURES	Patent: US 5723594-A 10 03-MAR-1998; Location/Qualifiers 1..86
BASE COUNT	22 a 21 c 25 g 18 t
ORIGIN	/organism="unknown"
Query Match	74.1%; Score 12.6; DB 5; Length 86;
Best Local Similarity	70.6%; Pred. No. 3.3e+03;
Matches 12; Conservative	2; Mismatches 3; Indels 0; Gaps 0;
OY	1 tgytngcrtcnacytg 17       :      :
Db	50 TGCTTAGCGTCACACTTG 66
RESULT 5	
LOCUS	E07909 48 bp DNA PAT 29-SEP-1997
DEFINITION	PCR primer.
ACCESSION	E07909
VERSION	E07909.1 GI:2176041
KEYWORDS	JP 1994205691-A/18.
SOURCE	unidentified.
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 48)
AUTHORS	Hayaishi,N.
TITLE	PRODUCTION OF MONOCLONAL ANTIBODY
JOURNAL	Patent: JP 1994205691-A 18 26-JUL-1994;
COMMENT	SANADELLI TECHNOL:KK; HAYASHI NAKANOBU OS None
OC	Artificial sequences.
PN	JP 1994205691-A/18
PD	26-JUL-1994
PF	08-JAN-1993 JP 1993001831
PI	HAYASHI NAKANOBU
PC	C12P1/08,C12N15/62//C12M15/13,(C12P21/08,C12R1:19); CC
strandedness:	Single;
CC	topology: Linear;



CC hypothetical: No; Location/Qualifiers  
FH Key  
FT source 1..48  
Location/Qualifiers  
source 1..48  
/organism="Artificial sequences"  
/db\_xref="taxon:32644"BASE COUNT 16 a 7 c 14 g 11 t  
ORIGIN

Query Match 71.8%; Score 12.2; DB 5; Length 48;  
Best Local Similarity 64.7%; Pred. No. 5.8e+03;  
Matches 11; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 tgytngcrtcnacytg 17  
11:11 1:11 1:11  
Db 37 TGTATTAGCTCACTG 21

RESULT 6  
LOCUS A59486 98 bp DNA PAT 06-MAR-1998  
DEFINITION Sequence 36 from Patent WO9705234.  
ACCESSION A59486  
VERSION A59486.1 GI:3714798  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified

REFERENCE 1 (bases 1 to 98)  
AUTHORS Chamberlain,S., Pook,M.A., Doudney,C., William,E., Hillermann,R.,  
TITLE Garcia-Valdecasas,J.J. and C.  
JOURNAL GENE FOR FRIEDREICH'S ATAXIA  
IMPERIAL COLLEGE (GB)  
FEATURES  
Source 1..98  
/organism="unidentified"  
/db\_xref="taxon:32644"BASE COUNT 30 a 23 c 22 g 23 t  
ORIGIN

Query Match 71.8%; Score 12.2; DB 5; Length 98;  
Best Local Similarity 64.7%; Pred. No. 5.8e+03;  
Matches 11; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 tgytngcrtcnacytg 17  
11:11 1:11 1:11  
Db 70 TGTTCGCGCTCCTG 54

RESULT 7  
LOCUS I42237 50 bp DNA PAT 07-OCT-1997  
DEFINITION Sequence 50 from patent US 5629153.  
ACCESSION I42237  
VERSION I42237.1 GI:2467732  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 50)  
AUTHORS Urdea,M.S.  
TITLE Use of DNA-dependent RNA polymerase transcripts as reporter  
molecules for signal amplification in nucleic acid hybridization  
assays  
JOURNAL Patent: US 5629153-A 50 13-MAY-1997;  
FEATURES  
Source Location/Qualifiers  
1..50

BASE COUNT 8 a 11 c 13 g 18 t  
ORIGIN

Query Match 65.9%; Score 11.2; DB 5; Length 50;  
Best Local Similarity 62.5%; Pred. No. 2.4e+04;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 tgytngcrtcnacyt 16  
11:11 1:11 1:1  
Db 21 TGTATGCACTCTCT 36

RESULT 8  
LOCUS AR051504/c 53 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 74 from patent US 5830670.  
ACCESSION AR051504  
VERSION AR051504.1 GI:5974868  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 53)  
AUTHORS de la Monte,S. and Wands,J.R.  
TITLE Neural thread protein gene expression and detection of Alzheimer's  
disease  
JOURNAL Patent: US 5830670-A 74 03-NOV-1998;  
FEATURES  
Source 1..53  
/organism="unknown"BASE COUNT 13 a 11 c 13 g 16 t  
ORIGIN

Query Match 65.9%; Score 11.2; DB 5; Length 53;  
Best Local Similarity 62.5%; Pred. No. 2.4e+04;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 gytngcrtcnacytg 17  
1:1 1:1 1:1 1:1  
Db 32 GCTATGCACTCACTG 17

RESULT 9  
LOCUS A43617 54 bp DNA PAT 06-MAR-1997  
DEFINITION Sequence 7 from Patent WO9507987.  
ACCESSION A43617  
VERSION A43617.1 GI:2298819  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified

REFERENCE 1 (bases 1 to 54)  
AUTHORS Budiol,M., De,W.C., Colau,D. and Roux-Salambien,P.  
TITLE NOVEL PROTEINS/POLYPEPTIDES AND COTRANSCRIPTION PLASMIDS AND LIVE  
RECOMBINANT CARRIERS THEREFOR  
JOURNAL Patent: WO 9507987-A 7 23-MAR-1995;  
COMMENT SOLVAY (BE)  
Other publication NO 961086 960509  
Other publication ZA 940687 950627  
Other publication AU 7615894 950403  
Other publication GB 2282601 950412.  
FEATURES  
Source 1..54  
/organism="unidentified"  
/db\_xref="taxon:32644"BASE COUNT 16 a 8 c 15 g 15 t  
ORIGIN

Query Match 65.9%; Score 11.2; DB 5; Length 54;  
Best Local Similarity 62.5%; Pred. No. 2.4e+04;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 tgyttngcrtcnacyt 17  
11:1 11:1 11:1 11:1  
Db 34 GTTGTGCGATCCACTG 19

RESULT 10  
AR040961 54 bp DNA PAT 29-SEP-1999  
LOCUS AR040961 Sequence 60 from patent US 5811238.  
DEFINITION AR040961  
ACCESSION AR040961  
VERSION AR040961.1 GI:5961457  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Stemmer, W.P.C. and Cramerl, A.  
TITLE Methods for generating polynucleotides having desired  
characteristics by iterative selection and recombination  
JOURNAL Patent: US 5811238-A 60 22-SEP-1998;  
FEATURES Location/Qualifiers  
source 1. .54  
BASE COUNT 9 a 18 c 15 g 12 t  
ORIGIN

Query Match 65.9%; Score 11.2; DB 5; Length 54;  
Best Local Similarity 62.5%; Pred. No. 2.4e+04;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 tgyttngcrtcnacyt 16  
11:1 11:1 11:1 11:1  
Db 7 TGCTGCGCATCCACCT 22

RESULT 11  
AR040962 57 bp DNA PAT 29-SEP-1999  
LOCUS AR040962 Sequence 61 from patent US 5811238.  
DEFINITION AR040962  
ACCESSION AR040962  
VERSION AR040962.1 GI:5961458  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 57)  
AUTHORS Stemmer, W.P.C. and Cramerl, A.  
TITLE Methods for generating polynucleotides having desired  
characteristics by iterative selection and recombination  
JOURNAL Patent: US 5811238-A 61 22-SEP-1998;  
FEATURES Location/Qualifiers  
source 1. .57  
BASE COUNT 13 a 17 c 10 g 17 t  
ORIGIN

Query Match 65.9%; Score 11.2; DB 5; Length 57;  
Best Local Similarity 62.5%; Pred. No. 2.4e+04;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 tgyttngcrtcnacyt 16  
11:1 11:1 11:1 11:1  
Db 7 TGCTGCGCATCCACCT 22

RESULT 12  
AR051866

LOCUS AR051866 57 bp DNA PAT 29-SEP-1999  
DEFINITION AR051866 Sequence 58 from patent US 5830721.  
ACCESSION AR051866  
VERSION AR051866.1 GI:5975230  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 57)  
AUTHORS Stemmer, W.P.C. and Cramerl, A.  
TITLE DNA mutagenesis by random fragmentation and reassembly  
JOURNAL Patent: US 5830721-A 58 03-NOV-1998;  
FEATURES Location/Qualifiers  
source 1. .57  
BASE COUNT 9 a 18 c 15 g 12 t 3 others  
ORIGIN

Query Match 65.9%; Score 11.2; DB 5; Length 57;  
Best Local Similarity 62.5%; Pred. No. 2.4e+04;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 tgyttngcrtcnacyt 16  
11:1 11:1 11:1 11:1  
Db 7 TGCTGCGCATCCACCT 22

RESULT 13  
AR051867 57 bp DNA PAT 29-SEP-1999  
LOCUS AR051867 Sequence 59 from patent US 5830721.  
DEFINITION AR051867  
ACCESSION AR051867  
VERSION AR051867.1 GI:5975231  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 57)  
AUTHORS Stemmer, W.P.C. and Cramerl, A.  
TITLE DNA mutagenesis by random fragmentation and reassembly  
JOURNAL Patent: US 5830721-A 59 03-NOV-1998;  
FEATURES Location/Qualifiers  
source 1. .57  
BASE COUNT 13 a 17 c 10 g 17 t  
ORIGIN

Query Match 65.9%; Score 11.2; DB 5; Length 57;  
Best Local Similarity 62.5%; Pred. No. 2.4e+04;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 tgyttngcrtcnacyt 16  
11:1 11:1 11:1 11:1  
Db 7 TGCTGCGCATCCACCT 22

RESULT 14  
AT5ABRE08 73 bp DNA PLN 21-MAR-1995  
LOCUS AT5ABRE08 Arabidopsis thaliana SABRE gene, exon 8.  
DEFINITION U19128  
ACCESSION U19128  
VERSION U19128.1 GI:719283  
KEYWORDS  
SEGMENT 8 of 14  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
Arabidopsis.  
REFERENCE 1 (bases 1 to 73)

AUTHORS Aeschbacher, R.A., Hauser, M.T., Feldmann, K.A. and Benfey, P.N.  
TITLE The SABRE gene is required for normal cell expansion in Arabidopsis  
JOURNAL Genes Dev. 9 (3), 330-340 (1995)  
MEDLINE 951/2383  
REFERENCE 2 (bases 1 to 73)  
AUTHORS Benfey, P.N.  
TITLE Direct Submision  
JOURNAL Submitted (27-DEC-1994) Phillip N. Benfey, Biology, New York  
University, 1009 Main Building, 100 Washington Square East, New  
York, NY 10003, USA

## FEATURES

source  
1..73  
/organism="Arabidopsis thaliana"  
/isolate="WS ecotype"  
/db\_xref="taxon:3702"  
/clone="SABRE"  
/chromosome="1"  
1..73  
/gene="SABRE"  
/number=8  
BASE COUNT 21 a 17 c 18 g 17 t  
ORIGIN

Query Match 65.9%; Score 11.2; DB 8; Length 73;  
Best Local Similarity 71.4%; Pred. No. 2.4e+04;  
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 tgytngcrtcnac 14  
||:| | | | | | | | | |  
Db 30 TGCTTGCAATCCAC 17

## RESULT 15

LOCUS CATFES13 84 bp DNA MAM 27-APR-1993  
DEFINITION Feline c-fes/fps proto-oncogene-encoded protein, exon 13.  
ACCESSION M16699  
VERSION M16699.1 GI:163840  
KEYWORDS fes/fps cellular oncogene; proto-oncogene.  
SEGMENT 13 of 19  
SOURCE Feline DNA.  
ORGANISM Felis catus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Carnivora; Fissipedia; Felidae; Felis.

REFERENCE 1 (bases 1 to 84)  
AUTHORS Roebroek, A.J.M., Schalken, J.A., Onnekink, C., Bloemers, H.P.J. and  
Van de Ven, W.J.  
TITLE Structure of the feline c-fes/fps proto-oncogene: Genesis of a  
retroviral oncogene

JOURNAL J. Virol. 61, 2009-2016 (1987)  
MEDLINE 87198954

## FEATURES

source  
1..84  
Location/Qualifiers  
/organism="Felis catus"  
/db\_xref="taxon:9685"  
prim\_transcript <1..>84  
intron <1..>20  
/note="c-fes/fps proto-oncogene mRNA and Intron"  
exon 21..74  
/number=13  
intron 75..>84  
/note="c-fes/fps Intron M"  
BASE COUNT 17 a 23 c 29 g 15 t  
ORIGIN About 360 bp after segment 12.

Query Match 65.9%; Score 11.2; DB 3; Length 84;  
Best Local Similarity 62.5%; Pred. No. 2.4e+04;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 gytngcrtcnac 17

Db 38 GTTAGCACCCTTG 23  
|:| | | | | | | | | |

Search completed: June 4, 2000, 16:05:26  
Job time: 27878 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:23:57 ; Search time 322.35 Seconds

(without alignments)  
13.195 Million cell updates/sec

Title: US-09-164-714-3

Perfect score: 17

Sequence: 1 tgytngcrtnacytg 17

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 431286

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12.6	74.1	86	1	T65226 Platelet derived g
2	12.6	74.1	86	1	T88603 VEGF nucleic acid
3	12.2	71.8	31	1	O90000 Modified Derf II m
4	12.2	71.8	98	1	T84384 Friedreich's ataxi
5	11.8	69.4	34	1	O90002 Modified Derf II m
6	11.2	65.9	50	1	O12936 Analyte capture pr
7	11.2	65.9	50	1	T74386 HIV analyte captur
8	11.2	65.9	54	1	O97417 Fragment of wild c
9	11.2	65.9	54	1	V54949 lacZ alpha gene st
10	11.2	65.9	57	1	O97418 Fragment of mutate
11	11.2	65.9	57	1	V54950 lacZ alpha gene st
12	11.2	65.9	74	1	T75059 Saponaria officina
13	11.2	65.9	31	1	O99185 Arabidopsis thalia
14	10.8	63.5	22	1	V21071 Arabidopsis RAP2.4
15	10.6	62.4	20	1	X16628 Interleukin 1 (441
16	10.6	62.4	22	1	X10014 Human biallelic po
17	10.6	62.4	24	1	O21863 Primer B, to demon
18	10.6	62.4	27	1	O24878 Probe E3-89 to hum
19	10.6	62.4	29	1	O89996 Modified Derf II m
20	10.6	62.4	32	1	V45582 Chicken alpha-amyl
21	10.6	62.4	36	1	O23643 JEV E positive str
22	10.6	62.4	43	1	V84742 Antibody heavy cha
23	10.6	62.4	45	1	O94185 Transmembrane olig
24	10.6	62.4	48	1	O76356 Antisense oligomer
25	10.6	62.4	50	1	O76357 Antisense oligomer
26	10.6	62.4	50	1	V37330 Antisense oligomer
27	10.6	62.4	62	1	V37333 ChimERIC Vpr gener
28	10.6	62.4	63	1	O41561 Oligonucleotide E.
29	10.6	62.4	65	1	N20038 Partial RNA sequen
30	10.6	62.4	75	1	O24882 Probe 811 to human
31	10.6	62.4	80	1	O46873 Ovine GHG primer O
32	10.6	62.4	87	1	O68700 Bacillus deramific
33	10.6	62.4	90	1	N92126 Sequence of Chalmy
34	10.6	62.4	97	1	T57696 Wheat germ aggluti

C 35	10.6	62.4	100	1	T30882 Primer 12 for 95 k
C 36	10.2	60.0	18	1	O28471 Thrombin aptamer p
C 37	10.2	60.0	18	1	T93171 5'-terminal flanki
C 38	10.2	60.0	18	1	V18757 Primer LE407 for c
C 39	10.2	60.0	18	1	V60398 5' PCR primer used
C 40	10.2	60.0	20	1	O82518 Chromosome 11 (loc
C 41	10.2	60.0	20	1	V03256 Homo sapiens mutan
C 42	10.2	60.0	21	1	X00302 LacZ specific PCR
C 43	10.2	60.0	22	1	T99764 Bacillus thuringie
C 44	10.2	60.0	22	1	T99735 Bacillus thuringie
C 45	10.2	60.0	25	1	V03257 Homo sapiens mutan

## ALIGNMENTS

RESULT 1	
T65226	
ID	T65226 standard; DNA; 86 BP.
AC	T65226;
DT	10-SEP-1997 (first entry)
DE	Platelet derived growth factor binding ligand 8.
KM	Platelet derived growth factor; PDGF; binding ligand;
KM	Identification; SELEX; anti-mitogenic; inhibition; cell;
KM	Systematic Evolution of Ligands by EXponential enrichment;
KM	epithelial; proliferation; diagnosis; treatment;
KM	neointimal lesion; arterial injury; ss.
OS	Synthetic.
PN	MO9638579-A1.
PD	05-DEC-1996.
PF	30-MAY-1996; U08014.
PR	02-JUN-1995; US-458424.
PR	02-JUN-1995; US-458423.
PR	05-JUN-1995; US-465594.
PR	05-JUN-1995; US-465591.
PR	07-JUN-1995; US-479783.
PR	07-JUN-1995; US-479725.
PR	20-MAR-1996; US-618693.
PA	(NEXS-) NEXSTAR PHARM INC.
PI	Gold L, Janjic N, Pagratlis N, Ringquist S, Toothman PJ;
DR	WPI; 97-034387/03.
PT	Identification of nucleic acid ligands to TGF-beta, PDGF and bKGF -
PT	using SELEX, used in the diagnosis and treatment of proliferative
PT	disorders
PS	Claim 39; Page 127; 209pp; English.
CC	The present sequence, a platelet derived growth factor (PDGF)
CC	binding ligand, was identified by Systematic Evolution of Ligands
CC	by EXponential enrichment (SELEX). Briefly a candidate mixture of
CC	nucleic acids was contacted with PDGF, and nucleic acids having an
CC	increased affinity to PDGF partitioned from the remainder of the
CC	mixture. The partitioned nucleic acids were then amplified to yield
CC	a mixture of nucleic acids enriched for sequences with higher
CC	affinity and specificity for binding to PDGF. The ligand is
CC	anti-mitogenic and may be used to inhibit epithelial cell
CC	proliferation, or in the diagnosis and treatment of PDGF mediated
CC	pathological conditions, e.g. to inhibit the development of
CC	neointimal lesions following arterial injury.
SQ	Sequence 86 BP; 22 A; 21 C; 25 G; 18 T;
Query Match	74.1%; Score 12.6; DB 1; Length 86;
Best Local Similarity	70.6%; Pred. No. 1.9e+02;
Matches 12; Conservative	2; Mismatches 3; Indels 0; Caps 0;
QY	1 tgytngcrtnacytg 17
DB	50 TGGTACGCTCACTTG 66
RESULT 2	
T88603	
ID	T88603 standard; DNA; 86 BP.
AC	T88603;

DT 18-JAN-1998 (first entry)  
 DE VEGF nucleic acid ligand used to treat angiogenesis-mediated diseases.  
 KM Nucleic acid ligand; secretory phospholipase A2; inflammation;  
 KM anti-inflammatory; rheumatoid arthritis; septic shock; VEGF;  
 KM vascular endothelial growth factor; angiogenesis; psoriasis;  
 KM metastasis; neoplasia; haemangioma; human immunodeficiency virus;  
 KM HIV; diagnosis; treatment; specific; high affinity; ss.  
 OS Synthetic.  
 PN WO9627604-A1.  
 PD 12-SEP-1996.  
 PF 04-MAR-1996: U02942.  
 PR 19-MAY-1995: US-447172.  
 PR 06-MAR-1995: US-399412.  
 PR 19-MAY-1995: US-447169.  
 PA (NEXS-) NEXSTAR PHARM INC.  
 PI Gold L, Janjic N, Lochrie M, Parma D;  
 DR WPI: 96-425373/42.  
 PT New non-naturally occurring nucleic acid ligands to secretory  
 PT phospholipase A2, VEGF and HIV-1 gag - useful e.g. for diagnosis  
 PT and treatment of inflammation, angiogenesis and HIV infection.  
 PS Claim 16; Page 71; 169pp; English.  
 CC T88594-T88671 are non-naturally occurring nucleic acid ligands (NAL)  
 CC for vascular endothelial growth factor (VEGF). Other non-naturally  
 CC occurring ligands for secretory phospholipase A2 (sPLA2) and HIV-1 gag  
 CC were also made. The NAL were identified using the SELEX procedure.  
 CC They are used as diagnostic or treatment agents. For example, those for  
 CC sPLA2 are used to study/diagnose inflammation and as anti-inflammatory  
 CC agents e.g. for the treatment of rheumatoid arthritis and septic shock;  
 CC those for VEGF are used to treat angiogenesis-mediated diseases e.g.  
 CC psoriasis, haemangioma, metastasis; and those for gag are used to  
 CC inhibit HIV-1 replication. NAL can also generally be used to identify  
 CC the corresponding target.  
 SQ Sequence 86 BP; 21 A; 22 C; 24 G; 18 T;

Query Match 74.1%; Score 12.6; DB 1; Length 86;  
 Best Local Similarity 70.6%; Pred. No. 1.9e+02;  
 Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 tgytngcrtcnacytg 17  
 ||||| ||||| |||||  
 DB 50 TGGTTCGCTCCACTTG 66

RESULT 3  
 ID 090000/c  
 AC 090000 standard; DNA: 31 BP.  
 DT 16-JAN-1996 (first entry)  
 DE Modified Derf II major tick allergen V5A forward primer.  
 KM Major tick allergen; Derf II; modified; mutein; anaphylactic shock;  
 KM mutagenic polymerase chain reaction primer; ss.  
 OS Synthetic.  
 PN J07095887-A.  
 PD 11-APR-1995.  
 PF 29-SEP-1993: 275897.  
 PR 29-SEP-1993: JP-275897.  
 PA (ASAK) ASAKI BREWERIES LTD.  
 PA (NIKK-) NIKKA WHISKEY KK.  
 PA (TORI) TORII YAKUHIN KK.  
 DR WPI: 95-174791/23.  
 PT Modified Derf II tick allergen with Ala subst. at specific position  
 PT - and its preparation by culturing transformant host cell, useful  
 PT for treating anaphylactic shock  
 PS Example 1: Page 5; 50pp; Japanese.  
 CC DerfII major tick allergen sequences which are modified versions of  
 CC three wild-type sequences are claimed (see R74567-R74569, along with  
 CC their corresp. cDNA coding sequences Q89963-Q89965 for the unmodified  
 CC allergens); the modified versions of these allergens have one  
 CC specified amino acid residue replaced by an Ala residue and are useful  
 CC for treating anaphylactic shock. See Q89966-Q89994 and R74570-R74598  
 CC for the 29 specifically claimed examples of modified DerfII major tick  
 CC allergens. The amino acid substitutions were introduced by mutation

CC of the corresp. codon in the wild-type coding sequence using PCR  
 CC primers Q89996-Q90016.  
 SQ Sequence 31 BP; 10 A; 4 C; 9 G; 8 T;

Query Match 71.8%; Score 12.2; DB 1; Length 31;  
 Best Local Similarity 64.7%; Pred. No. 2.8e+02;  
 Matches 11; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 tgytngcrtcnacytg 17  
 ||||| ||||| |||||  
 DB 28 TCTTTAGCATCGACTTG 12

RESULT 4  
 ID T84384/c  
 AC T84384 standard; DNA: 98 BP.  
 DT 12-NOV-1997 (first entry)  
 DE Friedreich's ataxia STM7 gene exon 21.  
 KM STM7 gene; Friedreich's ataxia; FRDA; neurodegeneration;  
 KM phosphatidylinositol-4-phosphate 5-kinase; PtdInsPK;  
 KM transgenic animal; animal model; diagnosis; therapy;  
 KM STM7.IIla; STM7.IIId; STM7.IIIc; ss.  
 OS Homo sapiens.  
 PN WO9705234-A2.  
 PD 13-FEB-1997.  
 PF 24-JUL-1996: G01786.  
 PR 28-JUL-1995: GB-015508.  
 PR 26-JUL-1995: GB-015309.  
 PA (UNLO) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.  
 PI Chamberlain S, Doudney CWF, Garcia-Valdecasas JJC;  
 PI Hillermann R, Pook MA;  
 DR WPI: 97-145674/13.  
 PT New gene, STM7, and new protein causing Friedreich's Ataxia -  
 PT encodes phosphatidylinositol phosphate kinase isoform; also splice  
 PT variants and transgenic animals, for diagnosing pre-disposition to  
 PT Friedreich's ataxia  
 PS Claim 1; Fig 16; 82pp; English.  
 CC DNA sequences (T84364-86) are provided for exons 1-23,  
 CC respectively, of the human STM7 gene associated with Friedreich's  
 CC ataxia (FRDA). The STM7.I gene (see T84358) comprises exons  
 CC 1-16 of the gene, while splice variant STM7.IIIa includes exons  
 CC 1-13, 17 and 19-22, splice variant STM7.IIId includes exons  
 CC 1-15, 17, 19, 21-22 and splice variant STM7.IIIc includes exons  
 CC 1-15 and 19-22. STM7 nucleic acids can be used to produce STM7  
 CC polypeptides (see W00976) in transfected host cells, to design  
 CC probes (esp. oligonucleotides from exons 1-16, partic. 7-11) used  
 CC in a claimed method for determining an inherited predisposition to  
 CC FRDA, to generate transgenic animal models of FRDA and in somatic  
 CC cell therapy.  
 SQ Sequence 98 BP; 30 A; 23 C; 22 G; 23 T;

Query Match 71.8%; Score 12.2; DB 1; Length 98;  
 Best Local Similarity 64.7%; Pred. No. 3.2e+02;  
 Matches 11; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 tgytngcrtcnacytg 17  
 ||||| ||||| |||||  
 DB 70 TGGTTCGCTCCTCTTG 54

RESULT 5  
 ID 090002/c  
 AC 090002 standard; DNA: 34 BP.  
 DT 16-JAN-1996 (first entry)  
 DE Modified Derf II major tick allergen D7A forward primer.  
 KM Major tick allergen; Derf II; modified; mutein; anaphylactic shock;  
 KM mutagenic polymerase chain reaction primer; ss.  
 OS Synthetic.  
 PN J07095887-A.

```

PD 11-APR-1995.
PF 29-SEP-1993: 275897.
PA (ASAK ) ASahi BREWERIES LTD.
PA (NIKK-) NIKKA WHISKY KK.
PA (TORI ) TORII YAKUHI KK.
DR WPI: 95-174791/23.
PT Modified Der fti tick allergen with Ala substd. at specific position
PT - and its preparation by culturing transformant host cell, useful
PT for treating anaphylactic shock
PS Example 1; Page 5; 50pp; Japanese.
CC derfrit major tick allergen sequences which are modified versions of
CC three wild-type sequences are claimed (see R74567-R74569, along with
CC their corresp. cDNA coding sequences Q89963-Q89965 for the unmodified
CC allergens); the modified versions of these allergens have one
CC specified amino acid residue replaced by an Ala residue and are useful
CC for treating anaphylactic shock. See Q89966-Q89994 and R74570-R74598
CC for the 29 specifically claimed examples of modified derfrit major tick
CC allergens. The amino acid substitutions were introduced by mutation
CC of the corresp. codon in the wild-type coding sequence using PCR
CC primers Q89996-Q90016.
SQ Sequence 34 BP; 9 A; 7 C; 9 G; 9 T;

Query Match 69.4%; Score 11.8; DB 1; Length 34;
Best Local Similarity 66.7%; Pred. No. 4.8e+02;
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 yttngcrtcnacyt 17
Db 26 TTTAGCATCTCTT 12

RESULT 6
Q12936 012936 standard; DNA: 50 BP.
ID 012936:
DT 11-OCT-1991 (first entry)
DE Analyte capture probe: HIV.97.XT1 for HIV.
KW Human Immunodeficiency Virus; ss.
OS Synthetic.
FH Key location/Qualifiers
FT misc-feature 1..30
FT /tag= a /note= "complementary to constant region of HIV "
FT /tag= b /note= "complementary to bound oligonucleotide"
FT misc-feature 31..50
FT /tag= b /note= "complementary to bound oligonucleotide"
PN MO9110746-A.
PD 25-JUL-1991.
PF 10-JAN-1991: U00213.
PR 10-JAN-1990; US-463022.
PA (CHIR-) CHIRON CORP.
PI Urdea MS:
DR WPI: 91-238034/32.
PT DNA-dependent RNA polymerase poly:deoxy:nucleotide construct -
PT used as signal amplifier in hybridisation assays, for detection
PT of N gonorrhoeae, HBV, HIV, etc.
PS Claim 69; Page 64; 88pp; English.
CC The analyte capture probe is used in a sandwich type hybridisation
CC assay in which a DNA-dependent RNA polymerase (T7) is used to
CC amplify the signal:
CC analyte capture template linker
CC ---w---|---x---|---a---
CC microtitre }---w---|---x---|---a---|---b---|---c---
CC dish } analyte template probe
CC The portion marked "y" hybridises with a specific portion of the
CC HIV genome and portion "w" hybridises with an oligonucleotide bound
CC to the microtitre dish. A template linker connects the analyte
CC with a T7 template probe which incorporates a T7 promoter and allows
CC amplification of the signal.
CC See also Q12914-Q12993.
SQ Sequence 50 BP; 8 A; 11 C; 13 G; 18 T;

```

```

Query Match 65.9%; Score 11.2; DB 1; Length 50;
Best Local Similarity 62.5%; Pred. No. 1.1e+03;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 tytgngcrtcnacyt 16
Db 21 TGTTCATCTCTCT 36

RESULT 7
T74386 74386 standard; DNA: 50 BP.
ID T74386:
DT 26-AUG-1997 (first entry)
DE HIV analyte capture probe: HIV.97.2.XT1.
KW Promoter; template probe; signal amplifier; hybridisation assay; detect;
KW functional domain; DNA-dependent RNA polymerase; quantifying; analyte;
KW ligand receptor; amplification; hepatitis B virus; Neisseria gonorrhoeae;
KW bacterial beta-lactamase TEM-1 gene; Chlamydia; HIV; hepatitis C virus;
KW bacterial tet M determinant; consensus; T7 promoter; ss.
OS Synthetic.
PN US5629153-A.
PD 13-MAY-1997.
PF 10-JAN-1990; 463022.
PR 10-JAN-1990; US-463022.
PR 10-JAN-1991; US-639560.
PR 08-MAR-1994; US-207901.
PA (CHIR ) CHIRON CORP.
PI Urdea MS:
DR WPI: 97-280266/25.
PT DNA construct for use as signal amplifier in hybridisation assays -
PT containing DNA-dependent RNA polymerase promoter and template
PT sequences
PS Example 9; Column 41-42; 45pp; English.
CC A novel DNA construct (referred to as a "template probe") for use as a
CC signal amplifier in hybridisation assays to detect a target comprises 3
CC functional domains (A, B and C) orientated A-B-C or B-C-A. (A) is single-
CC stranded and is designed to hybridise to complementary target sequence.
CC (B) is double-stranded and functions as a DNA-dependent RNA polymerase
CC promoter. (C) is single- or double-stranded, and functions as a template
CC for the promoter activity of domain B. It consists of a nucleotide
CC sequence not found in the template. The DNA construct is used in a method
CC for detecting and quantifying an oligonucleotide analyte or a ligand
CC receptor by amplification of a biological signal in a nucleic acid
CC hybridisation assay. The method is especially useful for determination of
CC nucleic acid segments characteristic of hepatitis B virus, Neisseria
CC gonorrhoeae, bacterial beta-lactamase TEM-1 gene, Chlamydia, bacterial
CC tet M determinant, HIV or hepatitis C virus. Comparison of template
CC probes with various numbers of base pairs between the A and B domain was
CC carried out by assaying for the presence of Human Immunodeficiency Virus
CC (HIV) DNA in human plasma. T74384-95 are specific analyte capture probes
CC used in the assay.
SQ Sequence 50 BP; 8 A; 11 C; 13 G; 18 T;

Query Match 65.9%; Score 11.2; DB 1; Length 50;
Best Local Similarity 62.5%; Pred. No. 1.1e+03;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 tytgngcrtcnacyt 16
Db 21 TGTTCATCTCTCT 36

RESULT 8
Q97417 097417 standard; DNA: 54 BP.
ID 097417:
AC Q97417;
DT 01-APR-1996 (first entry)
DE Fragment of wild type lacZ alpha gene.
KW lacZ alpha gene; recombination; chimERIC sequence; interleukin-1;

```

KM antibody; tPA; growth hormone; repetitive DNA; ribozyme; aptamer;  
KW gene therapy; ss.  
OS Synthetic.  
PN WO9522625-A1.  
PD 24-AUG-1995.  
PF 17-FEB-1995; U02126.  
PR 17-FEB-1994; US-198431.  
PI (AFFY-) AFFYMAX TECHNOLOGIES NV.  
PI Cramerl A, Stemmer WPC;  
DR WPI: 95-302727/39.  
PT DNA mutagenesis via random fragmentation and reassembly - useful in  
the prodn. of mutant proteins having enhanced biological activity  
PS Example 2; Fig 3; 120pp; English.  
CC The sequences shown in Q97417 and Q97418 are fragments of the LacZ alpha  
gene. Crossover between these two sequences was measured and indicated  
that gene shuffling was occurring. The sequence was reinserted into  
pUC18, and E.coli transformed with this plasmid were shown to have a  
lower frequency of LacZ resistance than by mutagenesis without shuffling.  
CC This method can also be used to shuffle whole gene sequences. The method  
can also be used to produce recombinant proteins, and chimeric DNA  
sequences. The methods can be used to produce proteins with a desired  
phenotype having an advantageous predetermined selectable property.  
CC Proteins such as interleukin-1, antibodies, tPA and growth hormones can  
be generated that have altered specificity or activity. Mutant sequences  
such as promoter regions, introns, exons and enhancer sequences can also  
be generated, which can be used to generate genes having increased  
expression rates. The method can be used in the study of repetitive DNA  
regions, and for the mutation of ribozymes or aptamers. Shuffled viral  
CC proteins which may comprise epitopes that will arise by viral evolution  
can also be produced. This method of shuffling sequences may be used for  
generating gene therapy vectors and replication-defective gene therapy  
constructs. The advantage with this method is that it is not necessary  
CC to know the actual DNA or RNA sequence of the specific nucleic acid  
fragment.  
SQ Sequence 54 BP; 9 A; 18 C; 15 G; 12 T;  
  
Query Match 65.9%; Score 11.2; DB 1; Length 54;  
Best Local Similarity 62.5%; Pred. No. 1.1e+03;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 tgyttingcrtcnacyt 16  
||:| ||:| ||:| ||:|  
Db 7 TGCTGGCGATCCACT 22  
  
RESULT 9  
V54949  
ID V54949 standard; DNA; 54 BP.  
AC V54949;  
DT 19-NOV-1998 (first entry)  
DE LacZ alpha gene stop codon mutant.  
KW LacZ alpha gene; recombination; gene shuffling; ss.  
OS Synthetic.  
PN US5811238-A.  
PD 22-SEP-1998.  
PF 30-NOV-1995; 564955.  
PR 04-MAR-1996; US-537874.  
PR 17-FEB-1994; US-198431.  
PR 30-NOV-1995; US-564955.  
PI (AFFY-) AFFYMAX TECHNOLOGIES NV.  
PI Cramerl A, Stemmer WPC;  
DR WPI: 98-530860/45.  
PT Production of polynucleotides with desired properties - by iterative  
selection and recombination  
PS Disclosure; Fig 3B; 74pp; English.  
CC V54947-50 represent fragments of LacZ alpha gene stop codon mutants,  
CC produced using the method of the invention. The specification  
CC describes a method for evolving a polynucleotide for acquisition  
CC of a desired property. The method comprises providing a population  
CC of variants of the polynucleotide, at least one of which is in  
CC cell-free form, shuffling the variants of the polynucleotide to form  
CC recombinant polynucleotides, selecting or screening for recombinant

CC polynucleotides that have evolved toward the desired property and  
CC repeating the steps with the selected recombinant polynucleotides  
CC until a recombinant polynucleotide has acquired the desired property.  
CC The recombinant polynucleotide produced by method, or its expression  
CC product, can be formulated as a pharmaceutical.  
SQ Sequence 54 BP; 9 A; 18 C; 15 G; 12 T;  
  
Query Match 65.9%; Score 11.2; DB 1; Length 54;  
Best Local Similarity 62.5%; Pred. No. 1.1e+03;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 tgyttingcrtcnacyt 16  
||:| ||:| ||:| ||:|  
Db 7 TGCTGGCGATCCACT 22  
  
RESULT 10  
Q97418  
ID Q97418 standard; DNA; 57 BP.  
AC Q97418;  
DT 01-APR-1996 (first entry)  
DE Fragment of mutated lacZ alpha gene.  
KW LacZ alpha gene; recombination; chimeric sequence; interleukin-1;  
KW antibody; tPA; growth hormone; repetitive DNA; ribozyme; aptamer;  
KW gene therapy; ss.  
OS Synthetic.  
PN WO9522625-A1.  
PD 24-AUG-1995.  
PF 17-FEB-1995; U02126.  
PR 17-FEB-1994; US-198431.  
PI (AFFY-) AFFYMAX TECHNOLOGIES NV.  
PI Cramerl A, Stemmer WPC;  
DR WPI: 95-302727/39.  
PT DNA mutagenesis via random fragmentation and reassembly - useful in  
the prodn. of mutant proteins having enhanced biological activity  
PS Example 2; Fig 3; 120pp; English.  
CC The sequences shown in Q97417 and Q97418 are fragments of the LacZ alpha  
gene. Crossover between these two sequences was measured and indicated  
CC that gene shuffling was occurring. The sequence was reinserted into  
CC pUC18, and E.coli transformed with this plasmid were shown to have a  
CC lower frequency of LacZ resistance than by mutagenesis without shuffling.  
CC This method can also be used to shuffle whole gene sequences. The method  
CC can also be used to produce recombinant proteins, and chimeric DNA  
CC sequences. The methods can be used to produce proteins with a desired  
CC phenotype having an advantageous predetermined selectable property.  
CC Proteins such as interleukin-1, antibodies, tPA and growth hormones can  
be generated that have altered specificity or activity. Mutant sequences  
such as promoter regions, introns, exons and enhancer sequences can also  
be generated, which can be used to generate genes having increased  
expression rates. The method can be used in the study of repetitive DNA  
CC regions, and for the mutation of ribozymes or aptamers. Shuffled viral  
CC proteins which may comprise epitopes that will arise by viral evolution  
can also be produced. This method of shuffling sequences may be used for  
CC generating gene therapy vectors and replication-defective gene therapy  
CC constructs. The advantage with this method is that it is not necessary  
CC to know the actual DNA or RNA sequence of the specific nucleic acid  
fragment.  
SQ Sequence 57 BP; 13 A; 17 C; 10 G; 17 T;  
  
Query Match 65.9%; Score 11.2; DB 1; Length 57;  
Best Local Similarity 62.5%; Pred. No. 1.1e+03;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 tgyttingcrtcnacyt 16  
||:| ||:| ||:| ||:|  
Db 7 TGCTGGCGATCCACT 22  
  
RESULT 11  
V54950  
ID V54950 standard; DNA; 57 BP.



AC V54950;  
DT 19-NOV-1998 (first entry)  
DE lacZ alpha gene stop codon mutant.  
KW lacZ alpha gene; recombination; gene shuffling; ss.  
OS Synthetic.  
PN US5811238-A.  
PD 22-SEP-1998.  
PF 30-NOV-1995; 564955.  
PR 04-MAR-1996; US-537874.  
PR 17-FEB-1994; US-198431.  
PR 30-NOV-1995; US-564955.  
PA (AFY-) AFYMAX TECHNOLOGIES NV.  
PI Cramer A, Stemmer WPC;  
DR WPI; 98-530860/45.  
PT Production of polynucleotides with desired properties - by iterative  
PT selection and recombination  
PS Disclosure; Fig 3B; 74pp; English.  
CC V54947-50 represent fragments of lacZ alpha gene stop codon mutants,  
CC produced using the method of the invention. The specification  
CC describes a method for evolving a polynucleotide for acquisition  
CC of a desired property. The method comprises providing a population  
CC of variants of the polynucleotide, at least one of which is in  
CC cell-free form, shuffling the variants of the polynucleotide to form  
CC recombinant polynucleotides, selecting or screening for recombinant  
CC polynucleotides that have evolved toward the desired property and  
CC repeating the steps with the selected recombinant polynucleotides  
CC until a recombinant polynucleotide has acquired the desired property.  
CC The recombinant polynucleotide produced by method, or its expression  
CC product, can be formulated as a pharmaceutical.  
SQ Sequence 57 BP; 13 A; 17 C; 10 G; 17 T;

Query Match 65.9%; Score 11.2; DB 1; Length 57;  
Best Local Similarity 62.5%; Pred. No. 1.1e+03;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 tgytngcrtcnacyt 16  
||:| ||:| ||:| ||:|  
Db 7 TGCTGCGATCGACCT 22

## RESULT 12

T75059  
ID T75059 standard; DNA; 74 BP.  
AC T75059;  
DT 04-SEP-1997 (first entry)  
DE Saponaria officinalis saporin gene C fragment sense primer.  
KW Saponaria officinalis; saporin; ribosome inactivating protein;  
KW targeting; internalisation; receptor binding internalised ligand;  
KW nucleic acid binding domain; cytoside; prodrug; prevention; RIP;  
KW epithelial; endothelial; keratinocyte; proliferation; eye;  
KW treatment; cancer; smooth muscle; hyperplasia; restenosis;  
KW primer; polymerase chain reaction; amplification; PCR; ss.  
OS Synthetic.  
PN WO9636362-A1.  
PD 21-NOV-1996.  
PF 16-MAY-1996; U07164.  
PR 16-MAY-1995; US-441979.  
PA (PRIZ-) PRIZM PHARM INC.  
PI Baird JA, Chandler LA, Sosnowski BA;  
DR WPI; 97-011855/01.  
PT Ligand and nucleic acid conjugates to treat excessive proliferation  
PT disorders - contain receptor binding ligand, nucleic acid binding  
PT domain and cytoside encoding agent or prodrug encoding agent  
PS Example 1; Page 89; 223pp; English.  
CC The present sequence is a primer for the PCR amplification of the  
CC C fragment of the gene encoding the Saponaria officinalis ribosome  
CC inactivating protein (RIP), saporin. The gene was used in the  
CC preparation of a novel pharmaceutical composition, comprising a  
CC receptor binding internalised ligand (RBIL), i.e. a polypeptide  
CC reactive with a cell surface receptor, nucleic acid binding domain  
CC (NABD) which binds to a nucleic acid and a cytoside encoding agent  
CC (CEA), i.e. a nucleic acid molecule encoding a cytoside or a

CC prodrug encoding agent (PEA), i.e. a nucleic acid encoding a  
CC prodrug, which is conjugated or fused in 1 of the following  
CC combinations RBIL-NABD-CEA, RBIL-NABD-PEA or RBIL-CEA-NABD and  
CC binds the cell surface receptor and internalises the CEA or PEA in  
CC cells bearing the receptor.  
CC The composition can be used in a medicament to prevent excessive  
CC epithelial cell, endothelial cell, fibroblast or keratinocyte  
CC proliferation in the eye. It can also be used to treat cancer, and  
CC smooth muscle cell hyperplasia. The addition of a linker increases  
CC the specificity, toxicity, solubility, serum stability or  
CC intracellular availability of the composition, and promotes nucleic  
CC acid condensation of the targeted moiety, while conjugates alter  
CC the function, gene expression or viability of a cell.  
SQ Sequence 74 BP; 15 A; 20 C; 24 G; 15 T;

Query Match 65.9%; Score 11.2; DB 1; Length 74;  
Best Local Similarity 62.5%; Pred. No. 1.1e+03;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 tgytngcrtcnacytg 17  
||:| ||:| ||:| ||:|  
Db 16 GCTGCGATCGACCTG 31

## RESULT 13

QY 099185/c  
ID 099185 standard; DNA; 31 BP.  
AC 099185;  
DT 07-MAR-1996 (first entry)  
DE Arabidopsis thaliana ecotype C24 trna Trp2-specific DNA primer TW1.  
KW Artificial male sterility; transgenic plant; barnase; toxin;  
KW crop improvement; ribonuclease; DNA primer; ss.  
OS Synthetic.  
PN WO9520668-A1.  
PD 03-AUG-1995.  
PF 31-JAN-1995; G00188.  
PR 31-JAN-1994; GB-001780.  
PA (GENE-) GENE SHEARS PTY LTD.  
PA (NICK-) NICKERSON BIOCHEM LTD.  
PI Betzner A, Huttner E, Lenee P, Paul W, Perez P;  
DR WPI; 95-275453/36.  
DT Prodn. of plants having a desired phenotypic trait - by crossing  
PT first and second lines which lack the trait, where one of the lines  
PT is transgenic  
PS Disclosure; Page 32; 78pp; English.  
CC The gene coding for trna Trp2 has been cloned from Arabidopsis  
CC thaliana ecotype C24, by PCR from genomic DNA, using primer TW1 and  
CC TW2 (099186).  
SQ Sequence 31 BP; 11 A; 7 C; 9 G; 4 T;

Query Match 64.7%; Score 11; DB 1; Length 31;  
Best Local Similarity 64.7%; Pred. No. 1.3e+03;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 tgytngcrtcnacyt 17  
||:| ||:| ||:| ||:|  
Db 26 TGCTTGCTTCCTCTCG 10

RESULT 14  
ID V21071  
V21071 standard; DNA; 22 BP.  
AC V21071;  
DT 25-AUG-1998 (first entry)  
DE Arabidopsis RAP2.4 (related to AP2) gene primer JORAP2.4L.  
KW APC; seed mass; AP2 domain containing gene; Arabidopsis; CMV; PCR;  
KW cauliflower mosaic virus; tobacco plant; primer; amplification; ss.  
OS Synthetic.  
PN WO9807842-A1.





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 13:52:56 ; Search time 4521.53 Seconds  
(without alignments)  
15.239 Million cell updates/sec

Title: US-09-164-714-3

Perfect score: 17  
Sequence: 1 tgyttingcrtcnacytg 17

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 119968

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: gb\_est1: \*  
21: gb\_est2: \*  
22: gb\_est3: \*  
23: gb\_est4: \*  
24: gb\_est5: \*  
25: gb\_est6: \*  
26: gb\_est7: \*  
27: gb\_est8: \*  
28: gb\_est9: \*  
29: gb\_est10: \*  
30: gb\_est11: \*  
31: gb\_est12: \*  
32: gb\_est13: \*  
33: gb\_est14: \*  
34: gb\_est15: \*  
35: gb\_est16: \*  
36: gb\_est17: \*  
37: gb\_est18: \*  
38: gb\_est19: \*  
39: gb\_est20: \*  
40: gb\_est21: \*  
41: gb\_est22: \*  
42: gb\_est23: \*  
43: gb\_est24: \*  
44: gb\_est25: \*

45: gb\_est26: \*  
46: gb\_est27: \*  
47: gb\_est28: \*  
48: gb\_est29: \*  
49: gb\_est30: \*  
50: gb\_est31: \*  
51: gb\_est32: \*  
52: em\_est20: \*  
53: em\_est21: \*  
54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*  
59: gb\_est33: \*  
60: gb\_est34: \*  
61: gb\_est35: \*  
62: gb\_est36: \*  
63: gb\_est37: \*  
64: gb\_est38: \*  
65: em\_est27: \*  
66: em\_est28: \*  
67: em\_est29: \*  
68: em\_est30: \*  
69: gb\_est39: \*  
70: gb\_est40: \*  
71: gb\_est41: \*  
72: gb\_est42: \*  
73: gb\_est43: \*  
74: gb\_est44: \*  
75: em\_est31: \*  
76: em\_est32: \*  
77: em\_est33: \*  
78: em\_est34: \*  
79: gb\_est45: \*  
80: gb\_est46: \*  
81: gb\_est47: \*  
82: gb\_gss1: \*  
83: gb\_gss2: \*  
84: gb\_gss3: \*  
85: gb\_gss4: \*  
86: em\_gss1: \*  
87: em\_gss2: \*  
88: em\_gss3: \*  
89: em\_gss4: \*  
90: gb\_gss5: \*  
91: gb\_gss6: \*  
92: gb\_gss7: \*  
93: gb\_gss8: \*  
94: gb\_gss9: \*  
95: em\_gss5: \*  
96: em\_gss6: \*  
97: em\_gss7: \*  
98: em\_gss8: \*  
99: em\_gss9: \*  
100: em\_gss10: \*  
101: em\_gss11: \*  
102: gb\_gss10: \*  
103: gb\_gss11: \*  
104: em\_gss12: \*  
105: gb\_gss12: \*  
106: gb\_gss13: \*  
107: gb\_gss14: \*  
108: gb\_gss15: \*  
109: gb\_gss16: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result

Query

No.	Score	Match Length	DB	ID	Description
1	12.8	75.3	80	21	T83169
2	12.8	75.3	80	27	AA045928
3	12.2	71.8	81	23	H23812
4	12.2	71.8	82	23	H40781
5	11.8	69.4	80	84	B14303
6	11.8	69.4	92	41	A1054253
7	11.8	69.4	97	50	A1669737
8	11.6	68.2	61	34	AA482394
9	11.6	68.2	67	42	A1159297
10	11.2	65.9	63	45	A1338554
11	11.2	65.9	65	33	AA420410
12	11.2	65.9	94	63	A19398803
13	11.2	65.9	100	36	AA659018
14	11.1	64.7	55	40	AA911593
15	11.1	64.7	73	39	AA839154
16	11.1	64.7	79	40	AA922367
17	11.1	64.7	81	69	AA119807
18	11.1	64.7	100	94	AQ025853
19	10.8	63.5	72	24	U38119
20	10.6	62.4	53	34	AA481226
21	10.6	62.4	54	24	D18203
22	10.6	62.4	61	39	AA906850
23	10.6	62.4	70	22	H06004
24	10.6	62.4	70	43	A1216035
25	10.6	62.4	73	41	A1001677
26	10.6	62.4	73	41	AA578173
27	10.6	62.4	78	23	H49037
28	10.6	62.4	79	27	AA009424
29	10.6	62.4	79	63	A1971513
30	10.6	62.4	79	63	A1973702
31	10.6	62.4	83	20	T50051
32	10.6	62.4	86	60	AA244290
33	10.6	62.4	88	62	A18773394
34	10.6	62.4	92	24	H88263
35	10.6	62.4	93	43	A124211
36	10.6	62.4	94	34	AA487086
37	10.6	62.4	96	28	AA086143
38	10.6	62.4	99	20	T48764
39	10.6	62.4	99	38	AA745750
40	10.6	62.4	100	20	T51146
41	10.6	62.4	100	22	R23019
42	10.4	61.2	53	21	T58978
43	10.4	61.2	88	42	A1149323
44	10.2	60.0	41	82	HSK119703
45	10.2	60.0	46	40	AA953796
					T83169 yd41d05.r1
					AA045928 zk75c01.r
					H23812 yn70h11.s1
					H40781 yo08c10.s1
					B14303 A-345G4.TV
					A1054253 j15h505.x
					A1669737 tui2e08.x
					AA482394 t34a602.r
					A1159297 vz85e01.r
					A1338554 q933c08.x
					AA420410 vc51e07.r
					A19398803 sc60d07.y
					AA659018 n0179f09.s
					AA911593 c090c01.s
					AA839154 vW48c07.r
					AA922367 Oh91b11.s
					AA119807 sd53a03.y
					AQ025853 1(2)k0801
					U38119 OSU8119.FD
					AA481226 aa34f07.r
					D18203 MUGS00473
					AA906850 o120d04.s
					H06004 y176c06.r1
					A1216035 qn10e11.x
					A1001677 EST0259.T
					AA578173 n121g07.s
					H49037 EST00180.S
					AA009424 zeb2d01.r
					A1971513 w05f02.x
					A1973702 s008f12.y
					T50051 yb04b03.r1
					AA244290 nc06a06.r
					A18773394 vz74b07.r
					H88263 yw20h02.s1
					A124211 ch92a10.x
					AA487086 db24g01.r
					AA086143 18309.seq
					T48764 yb02f06.r1
					AA745750 ob02a02.s
					T51146 yb94d01.s1
					R23019 yb27d06.r1
					T58978 yb59c02.r1
					A1149323 qc76f09.x
					X88317 H.sapiens D
					AA953796 co038b09.s

## ALIGNMENTS

RESULT	1
LOCUS	T83169
DEFINITION	T83169 80 bp mRNA EST 16-MAR-1995 yddl05.r1 Soares fetal liver spleen INFUS Homo sapiens cDNA clone IMAGE:110793 5', mRNA sequence.
ACCESSION	T83169
VERSION	T83169.1 GI:711457
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo. 1 (bases 1 to 80)
AUTHORS	Hillier,L., Clark,N., Dubaque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
TITLE	The WashU-Merck EST Project
JOURNAL	Unpublished (1995)
COMMENT	Contact: Wilson RK

```

FEATURES
source
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Insert Size: 935
Source: IMAGE Consortium, LNLN. This clone is available royalty-free
through LNLN ; contact the IMAGE Consortium (info@image.lnl.n.gov)
for further information. Putative full length read
Insert Length: 935 Std Error: 0.00
Seq primer: M13RPI
High quality sequence stop: 251.
Location/Qualifiers
1. 80
/organism="Homo sapiens"
/db_xref="GDB:466410"
/db_xref="taxon:9606"
/clone="IMAGE:110793"
/clone_1b="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker. Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
15' AACGTGACAGATTAATTAAGAGCTTTTGTGTGTGTGTGT 3'1,
(double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT
23 a 19 c 11 g 27 t
ORIGIN

```

Query Match	75.3%	Score 12.8	DB 21	Length 80
Best Local Similarity	68.8%	Pred. No. 1	1e+03	
Matches	11	Conservative	3	Mismatches 2; Indels 0; Gaps 0;
QY	1	tgtgttngcrtcnacgt	16	
Db	4	TGTTTTCATCACTT	19	
RESULT	2			
AA045928				
LOCUS				
DEFINITION	AA045928	80 bp	mRNA	EST
ACCESSION	AA045928			11-MAY-1997
VERSION	AA045928.1	GI:1525822		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 (bases 1 to 80)			
	Hillier, L., Lennon, G., Becker, M., Donald, M.F., Chiappelli, B., Chissole, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R. and Warr, M.			
TITLE	Generation and analysis of 280,000 human expressed sequence tags			
JOURNAL	Genome Res. 6 (9), 807-828 (1996)			
MEDLINE	97044478			
COMMENT	On Apr 14, 1993 this sequence version replaced gi:716787. Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810			

Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
Putative full length read  
The vector to vector length is 81  
Insert Length: 822 Std Error: 0.00  
Seq primer: -28M13 rev2 from Amersham.

FEATURES  
source

1..80  
/organism="Homo sapiens"  
/db\_xref="GDB:3803227"  
/db\_xref="taxon:9606"  
/clone="IMAGE:488640"  
/clone\_lib="Soares\_adult brain N2B5HB55Y"  
/sex="Female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pT73-Pac; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGAGAGATTGCGCGCCCTTTTCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."

BASE COUNT  
ORIGIN

23 a 19 c 11 g 27 t

## Query Match

Best Local Similarity 75.3%; Score 12.8; DB 27; Length 80;

Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgytngcrtcnacyt 16  
||:| | | | | | | |

Db 4 TGTTCGATCGACTT 19

## RESULT 3

## LOCUS

H23812 61 bp mRNA EST 06-JUL-1995

DEFINITION Y070H11.s1 Soares adult brain N2B5HB55Y Homo sapiens cDNA clone

IMAGE:173829 3' similar to SP:K02F3.2 CE01348 ; mRNA sequence.

## ACCESSION

H23812 H23812.1 GI:892507

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

TITLE  
JOURNAL  
COMMENT

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 61)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)  
On Sep 21, 1992 this sequence version replaced gi:276165.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 1692  
High quality sequence starts: 1  
High quality sequence stops: 1  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Trace considered overall poor quality

Possible reversed clone: similarity on wrong strand  
Insert Length: 1692 Std Error: 0.00  
Seq primer: Promega -21m13  
High quality sequence stop: 1.

FEATURES  
source

1..61  
/organism="Homo sapiens"  
/db\_xref="GDB:3835864"  
/db\_xref="taxon:9606"  
/clone="IMAGE:173829"  
/clone\_lib="Soares\_adult brain N2B5HB55Y"  
/sex="Male"  
/dev\_stage="55-year old"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: brain; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTACCAATCTGAAGTGGAGCGCGCCGCTTTTCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 53. Library constructed by Bento  
Soares and M. Fatima Bonaldo. The adult brain RNA was  
provided by Dr. Donald H. Gilden. Tissue was acquired  
17-18 hours after death which occurred in consequence of a  
ruptured aortic aneurysm. RNA was prepared from a pool of  
tissues representing the following areas of the brain:  
frontal, parietal, temporal and occipital cortex from the  
left and right hemispheres, subcortical white matter,  
basal ganglia, thalamus, cerebellum, midbrain, pons and  
medulla."

BASE COUNT 14 a 21 c 16 g 10 t

## ORIGIN

## Query Match

Best Local Similarity 71.8%; Score 12.2; DB 23; Length 61;

Matches 11; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tgytngcrtcnacyt 17  
||:| | | | | | | |

Db 12 TGTTCGATCGACTG 28

## RESULT 4

## LOCUS

H40781 82 bp mRNA EST 31-JUL-1995

DEFINITION Y08C10.s1 Soares adult brain N2B5HB55Y Homo sapiens cDNA clone

IMAGE:177330 3' similar to SP:K02F3.2 CE01348 ; mRNA sequence.

## ACCESSION

H40781 H40781.1 GI:916833

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

TITLE  
JOURNAL  
COMMENT

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 82)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)  
On May 9, 1995 this sequence version replaced gi:803137.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 1926

```

High quality sequence starts: 1
High quality sequence stops: 1
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Trace considered overall poor quality
Possible reversed clone: similarly on wrong strand
Insert Length: 1926 Std Error: 0.00
Seq Primer: Promega -21ml13
High quality sequence stop: 1.

```

## FEATURES

```

SOURCE
1. 82
/organism="Homo sapiens"
/db_xref="GDB:3839526"
/db_xref="taxon:9606"
/clone="IMAGE:177330"
/clone_id="Soares adult brain N2B5HB55Y"
/sex="Male"
/dev_stage="55-year old"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: brain; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with Not I - oligo(dT) primer [5
TGTTACCAATCTGAAGTGGAGCGCGCGCTTTTCTTTTCTTTTCTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 53. Library constructed by Bento
Soares and M. Fatima Bonaldo. The adult brain RNA was
provided by Dr. Donald H. Gilden. Tissue was acquired
17-18 hours after death which occurred in consequence of a
ruptured aortic aneurysm. RNA was prepared from a pool of
tissues representing the following areas of the brain:
frontal, parietal, temporal and occipital cortex from the
left and right hemispheres, subcortical white matter,
basal ganglia, thalamus, cerebellum, midbrain, pons and
medulla."

```

BASE COUNT	17	a	28	c	23	g	13	t	1	others
ORIGIN										
Query Match	71.8%: Score 12.2; DB 23; Length 82;									
Best Local Similarity	64.7%: Pred. No. 2.5e+03;									
Matches 11: Conservative	3; Mismatches 3; Indels 0; Gaps 0;									
OY	1	tgtytngertcnacytg	17							
		:     :								
Db	33	TGTYTTCATCGACCTG	49							

RESULT	5
B14303	
LOCUS	B14303 80 bp DNA GSS 04-JUN-1998
DEFINITION	A-345G4.TV C19785KAI Homo sapiens genomic clone A-345G4, genomic survey sequence.
ACCESSION	B14303
VERSION	B14303.1 GI:2122052
KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Euarcharia; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 80) Adams,M.D., Kelley,J.M., Rounsley,S.R. and Venter,J.C. Use of a BAC End Sequence Database for Sequence-Ready Map Building Unpublished (1997)
TITLE	Other GSSs: 345G4.TP
JOURNAL	Contact: Mark Adams
COMMENT	Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200  
Fax: 301 838 0208  
Email: [mdadamset@ligr.org](mailto:mdadamset@ligr.org)  
Clones are available from Research Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC  
end search page:  
[http://www.ligr.org/tbdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.ligr.org/tbdb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: T7  
Class: BAC ends.

## FEATURES

Source	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Source	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Source	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Source	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Source	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Source	1	2	3	4	5	6																																																																																														

Query Match	69.4%	Score 11.8	DB 84	length 80
Best Local Similarity	62.5%	Pred. No. 4.3e+03		
Matches	10	Conservative	3	Mismatches 3
				Indels 0
				Gaps 0
QY	1	tgytngcrctcnac	16	
		:     :	:	
Db	12	tgcttgcgcanccact	27	

RESULT	6
LOCUS	AI054253
DEFINITION	AI054253 92 bp mRNA EST 27-JUL-1998
VERSION	AI054253.x1 NC1_CGAP_Ov26 Homo sapiens cDNA clone IMAGE:1862361 3'
KEYWORDS	mRNA sequence.
ORGANISM	AI054253
SOURCE	AI054253.1 GI:3322040
	EST.
	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
TITLE	Eutheria; Primates; Catarrhini; Hominiidae; Homo.
JOURNAL	1 (bases 1 to 92)
COMMENT	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
	Tumor Gene Index
	Unpublished (1997)
	On Jan 19, 1998 this sequence version replaced gi:2287056.

## FEATURES

```
source
1. .92
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1862361"
/clone_id="NCI-CGAP_Ov26"
/sex="Female"
/tissue_type="papillary serous carcinoma"
/dev_stage="adult"
/lab_host="DHI08"
```



Query match 69.48; Score 11.8; DB 50; Length 97;

Query Match	68.2%	Score 11.6	DB 34	length 61
Best Local Similarity	68.8%	Pred. No. 5	2e03	
Matches	11	Conservative	2	Mismatches 3
				Indels 0
				Gaps 0
Oy	2	gyltngertcncacg	17	
db	23	GATTAGCATCAACCTTG	8	

AI159297/c	AI159297	67 bp	mRNA	EST	02-OCT-1998
LOCUS					
DEFINITION	v255e01.1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone				
	IMAGE:1333272 5' similar to WP:MI53.1 CE03511				
	PYRROLINE-5-CARBOXYLATE REDUCTASE ;, mRNA sequence.				
ACCESSION	AI159297				
VERSION	AI159297.1 GI:3692479				
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 67)				
AUTHORS	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,				
	Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, D., Morris, M.,				
	Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,				
	Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and				
	Waterston, R.				
TITLE	The Mashu-HMI Mouse EST Project				
JOURNAL	Unpublished (1996)				
COMMENT	On Jan 19, 1998 this sequence version replaced gi:2287146.				
	Washu-HMI Mouse EST Project				
	Washington University School of Medicine				
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108				
	Tel: 314 286 1800				
	Fax: 314 286 1810				
	Email: mouse@watson.wustl.edu				
	This clone is available royalty-free through LLNL ; contact the				
	IMAGE Consortium (info@image.llnl.gov) for further information.				
	GI:692816				
	Trace considered overall poor quality				
	possible reversed clone; similarity on wrong strand				
	Seq primer: -28m3 rev2 ET from Amersham				
	High quality sequence stop: 1.				
FEATURES	Location/Qualifiers				
source	1..67				
	/organism="Mus musculus"				
	/strain="C57BL/6J"				
	/db_xref="taxon:10090"				
	/clone="IMAGE:1333272"				
	/clone_11b="Soares_mammary_gland_NbMMG"				
	/sex="male"				
	/tissue.type="mammary gland"				
	/dev_stage="4 weeks"				
	/lab_host="DH10B"				
	/note="Organ: mammary gland; Vector: pT7T3D-Pac				
	(Pharmacia) with a modified polylinker; Site:1: Not I;				
	Site:2: Eco RI; 1st strand cDNA was primed with a Not I -				
	oligo(dT) primer [5',				
	TGTTACCAATCTGCAAGTGGCGCGCGCAATGTTTTTTTTTTTTTTTTTTTTTTT				
	T 3']; double-stranded cDNA was ligated to Eco RI				
	adaptors (Pharmacia), digested with Not I and cloned into				
	the Not I and Eco RI sites of the modified pT7T3 vector.				
	RNA provided by Dr. Minoru Ko, Wayne State Univ. library				
	constructed and normalized by Bento Soares and M.Patlina				
	Bonafide."				
BASE COUNT	21 a	19 c	18 g	9 t	
ORIGIN					
Query Match	68.2%	Score 11.6;	DB 42;	Length 67;	
Best Local Similarity	68.8%	Pred. No. 5.4e+03;			
Matches 11; Conservative	2;	Mismatches 3;	Indels 0;	Gaps 0;	
Qy	1 tgyttingcrtcnacyt 16				
	:				
Db	48 TGTTTAGCTTCTACTT 33				
RESULT	10				

LOCUS	AI138554/c	EST	29-DEC-1998
DEFINITION	AI138554 63 bp mRNA		
	g993c08.x1 Soares_total_fetus.NB2HF8.9w Homo sapiens cDNA clone IMAGE:1938926 3' similar to gb:U02067 H.sapiens mRNA for 7SL RNA pseudogene (HUMANN); contains element MER30 repetitive element ; , mRNA sequence.		
ACCESSION	AI138554		
VERSION	AI138554.1 GI:4075481		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
REFERENCE	1 (bases 1 to 63)		
AUTHORS	NCI-CCAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT	On Feb 11, 1998 this sequence version replaced gi:2873097. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov This clone is available royalty-free through LNL; contact the IMAGE Consortium ( <a href="http://infoimage.lnl.gov">infoimage.lnl.gov</a> ) for further information. Trace considered overall poor quality Seq primer: -400p from Glbco High quality sequence stop: 1. Location/Qualifiers 1..63 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1938926" /clone_lib="Soares_total_fetus_NB2HF8_9w" /dev_stage="8-9 weeks" /lab_host="DH10B" /note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCTTAATTTTCTTTT 3'] . Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3D vector. Library went through one round of normalization, and was constructed by Benito Soares and W. Fatima Bonaldo. "		
BASE COUNT	18 a 21 c 12 g 12 t		
ORIGIN			
Query Match	65.9%; Score 11.2; DB 45; Length 63;		
Best Local Similarity	62.5%; Pred. No. 9.2e+03;		
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;			
QY	2 gytgtcrtcnacytg 17		
	:     :		
Db	49 GTTGGCAATCAATTG 34		
RESULT 11			
LOCUS	AA420410 65 bp mRNA	EST	16-OCT-1997
DEFINITION	vc51607.f1 Knowles Solter mouse 2 cell Mus musculus cDNA clone IMAGE:778116 3', mRNA sequence.		
ACCESSION	AA420410		
VERSION	AA420410.1 GI:2080907		
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 65)		
AUTHORS	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,		

TITLE  
The WashU-HMI Mouse EST Project  
JOURNAL  
Unpublished (1996)  
COMMENT  
On Sep 12, 1996 this sequence version replaced gi:1394455.

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through INL; contact the  
IMAGE Consortium ([info@image.jlnl.gov](mailto:info@image.jlnl.gov)) for further information.  
G1:470972

```

High quality sequence stop: 64.
FEATURES
location/Qualifiers
source
1. .65

```

```

/organism="Mus musculus"
/strain="B6D2 F1/J"
/db_xref="taxon:10090"
/clone_image="778116"
/clone_id="Knowles Solter mouse\2 cell"
/tissue_type="embryo"
/dev_stage="2-cell"
/lab_host="DH10B"
/note="Organ: embryo; Vector: pBluescribe (modified)
Site_1: Mu1; Site_2: SalI; Cloned unidirectionally;
mRNA prepared from 13,500 2-cell stage embryos.
SalI (dr): 5'-CGGCGACCGCGACCGCTTTTTTTTTT-3'."
were cloned into the Mu1/SalI sites of a modified
pBluescribe vector using commercial linkers (NEB).
Average insert size: 1.2 kb."

```

Query Match	65.9%	Score 11.2	DB 33	Length 65
Best Local Similarity	62.5%	Pred. No. 9.3e+03		
Matches 10	Conservative 3	Mismatches 3	Indels 0	Gaps 0
QY	2	gyttingcrtcmacytg	17	
		: :	:	:
Db	54	gtyttingcrtcmacytg	39	

RESULT	12
A1938803/c	
LOCUS	A1938803 94 bp mRNA EST 13-DEC-1999
DEFINITION	scc0d07.y1 Gm-c1016 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
ACCESSION	Gm-c1016-806 5' similar to TR:O04132 O04132 SRC1 ; , mRNA sequence.
VERSION	A1938803
KEYWORDS	A1938803.1 GI:5677673
SOURCE	EST.
ORGANISM	soybean. glycine max

REFERENCE	1 (bases 1 to 94)
AUTHORS	Shoemaker, R., Keim, P., Vockin, L., Expediting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Pearson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurr, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
TITLE	Public Soybean EST Project
JOURNAL	Unpublished (1999)
COMMENT	On Jun 5, 1998 this sequence version replaced gi:3189209. Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [estewartson.wustl.edu](mailto:estewartson.wustl.edu)

Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 fax: (888) 919-3324 or (314) 427-3324 or contact: [clones@genomesystems.com](mailto:clones@genomesystems.com) or [info.genomesystems.com](http://info.genomesystems.com) web site: [www.genomesystems.com](http://www.genomesystems.com)  
Trace considered overall poor quality  
Possible reversed clone: similarity on wrong strand  
High quality sequence stop: 1.

**FEATURES**  
**source**

```

BASE COUNT      33 a      23 c      23 g      15 t
ORIGIN
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl016-806"
/clone_lib="Gm-cl016"
/tissue_type="immature flowers of field grown plants"
/lab_host="XLI0-Gold"
/notice="Vector: plusscript II XR; Site.1: EcoRI; Site.2:
XhoI; This cDNA library was constructed from mRNA isolated
from immature flowers of field grown plants. The cDNA
library was prepared using the Stratagene pluscript II
XR library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dT) sequence with a XhoI restriction site. EcoRI adapters
were ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
plusscript vector. The ligated cDNA fragments were
transformed into XLI0-Gold host cells. This library was
constructed by Dr. Randy Shoemaker and Dr. John
Erpelding."

```

Query Match	65.9%	Score 11.2	DB 63	Length 94
Best Local Similarity	62.5%	Pred. No. 1e+04		
Matches 10; Conservative	3	Mismatches	3	Indels 0
QY 1 tgytngertcnaeyt 16				
:				
Db 90 tgyrttgcctcaacgt 75				

RESULT	13		
AA659018			
LOCUS	100 bp	RNA	EST
DEFINITION	nu79ff09.s1 NCI-GCAP Alvi Homo sapiens cDNA IMAGE:1216537 similar to SW:R27A_HUMAN P14798 40S RIBOSOMAL PROTEIN S27A. (1) ;		
ACCESSION	AA659018		
VERSION	AA659018		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 100)  
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On May 5, 1995 this sequence version replaced gi:797829.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,  
Ph.D.  
cDNA Library Preparation: David B. Krizman, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/dbcrp/image/image.html

Trace considered overall poor quality  
Insert Length: 376 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers

## FEATURES

source

```
1..100
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1216937"
/clone_lib="NCI-CGAP_Alvi"
/tissue_type="alveolar rhabdomyosarcoma"
/lab_host="DH10B"
/note="Vector: PAMPI0; mRNA made from alveolar
rhabdomyosarcoma, cDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel,
average insert size 600 bp. Reference: Krizman et al.
(1996) Cancer Research 56:5380-5383."
BASE COUNT      21 a      24 c      20 g      35 t
ORIGIN
```

Query Match 65.9%; Score 11.2; DB 36; Length 100;  
Best Local Similarity 71.4%; Pred. No. 1e+04; 2; Indels 0; Gaps 0;  
Matches 10; Conservative 2; Mismatches 2;

QY 4 ttngtcnacytg 17  
||| ||| ||| |||  
Db 12 TTCGATCCACCTG 25

## RESULT 14

AA911593

LOCUS AA911593 55 bp mRNA EST 21-APR-1998  
DEFINITION o90gc01.s1 NCI-CGAP Br5 Homo sapiens cDNA clone IMAGE:1377504  
Similar to TR:05157 055157 HYPOTHETICAL 23.6 KD PROTEIN. ; contains  
MER22 t3 MSRI repetitive element ;, mRNA sequence.

ACCESSION AA911593  
VERSION AA911593.1 GI:3050957  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 55)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)

JOURNAL On Jan 19, 1998 this sequence version replaced gi:2285536.  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Ilan Kirsch, M.D., Kristina A. Cole, M.D.,  
Ph.D. student, Rodrigo F. Chuaguí, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.  
cDNA Library Preparation: David B. Krizman, Ph.D.  
DNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/dbcrp/image/image.html

Trace considered overall poor quality  
Insert Length: 441 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers

## FEATURES

source

```
1..55
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1377504"
/clone_lib="NCI-CGAP_Br5"
/sex="female"
/tissue_type="infiltrating ductal carcinoma"
/lab_host="DH10B"
/note="Organ: breast; Vector: PAMPI0; mRNA made from
infiltrating ductal carcinoma, cDNA made by oligo-dT
priming. Non-directionally cloned. Size-selected on
agarose gel, average insert size 600 bp."
BASE COUNT      9 a      22 c      12 g      12 t
ORIGIN
```

Query Match 64.7%; Score 11; DB 40; Length 55;  
Best Local Similarity 64.7%; Pred. No. 1.2e+04; 4; Indels 0; Gaps 0;  
Matches 11; Conservative 2; Mismatches 4;

QY 1 tgytngtcnacytg 17  
||| ||| ||| |||  
Db 24 TGTCTGATCAACATG 40

## RESULT 15

AA839154

LOCUS AA839154 73 bp mRNA EST 27-FEB-1998  
DEFINITION v48c07.r1 Soares mammary-gland NBMGC Mus musculus cDNA clone  
IMAGE:1247052 5' similar to SW:STO\_HUMAN P47897 GLUTAMINYL-TRNA  
SYNTHETASE ;, mRNA sequence.

ACCESSION AA839154  
VERSION AA839154.1 GI:2915249  
KEYWORDS EST.

SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 73)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dierich,N., Dubuque,T.,  
Geisler,S., Knabbe,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Weisling,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
TITLE The WashU-HMI Mouse EST Project  
Unpublished (1996)

JOURNAL On Jan 19, 1998 this sequence version replaced gi:2150179.  
COMMENT Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:660740

Trace considered overall poor quality  
Possible reversed clone: similarity on wrong strand  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers

## FEATURES

source

```
1..73
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1247052"
/clone_lib="Soares_mammary_gland_NBMGC"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT73D-Pac
```

## FEATURES

BASE COUNT  
ORIGIN

Best Local  
Matches 1

Db 28 TCCT

Search complete  
Job time: 209

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:09:37 ; Search time 244.64 Seconds  
(without alignments)  
9.033 Million cell updates/sec

Title: US-09-164-714-3

Perfect score: 17  
Sequence: 1 tgytngctcnaacytg 17

Scoring table: IDENTITY\_NDC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 375880

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

1: Issued\_Patents\_NA:\*  
2: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5C.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/5D.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/6.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12.6	74.1	86	1	US-08-479-783A-10 Sequence 10, Appl
2	12.6	74.1	86	1	US-08-479-725-10 Sequence 10, Appl
3	12.6	74.1	86	2	US-08-618-693-10 Sequence 10, Appl
4	12.6	74.1	86	2	US-08-447-169A-168 Sequence 168, Appl
5	12.6	74.1	86	6	PCT-US96-08014-99 Sequence 99, Appl
6	11.2	65.9	27	4	US-08-859-998-1003 Sequence 1003, Appl
7	11.2	65.9	50	1	US-08-207-901-50 Sequence 50, Appl
8	11.2	65.9	53	3	US-08-454-557C-74 Sequence 74, Appl
9	11.2	65.9	53	4	US-08-340-426D-74 Sequence 74, Appl
10	11.2	65.9	53	4	US-08-450-673C-74 Sequence 74, Appl
11	11.2	65.9	53	6	PCT-US95-17111A-74 Sequence 74, Appl
12	11.2	65.9	54	2	US-08-564-955-60 Sequence 60, Appl
13	11.2	65.9	57	2	US-08-564-955-61 Sequence 61, Appl
14	11.2	65.9	57	3	US-08-537-874-58 Sequence 58, Appl
15	11.2	65.9	57	3	US-08-537-874-59 Sequence 59, Appl
16	11.2	65.9	74	5	US-08-718-904-97 Sequence 97, Appl
17	10.8	63.5	21	4	US-09-028-361A-7 Sequence 7, Appl
18	10.6	62.4	22	1	US-08-379-078-430 Sequence 430, Appl
19	10.6	62.4	22	1	US-08-379-078-431 Sequence 431, Appl
20	10.6	62.4	22	1	US-08-379-078-432 Sequence 432, Appl
21	10.6	62.4	27	1	US-07-766-351-9 Sequence 9, Appl
22	10.6	62.4	27	1	US-08-059-032-9 Sequence 9, Appl
23	10.6	62.4	27	6	PCT-US91-07290-9 Sequence 9, Appl
24	10.6	62.4	30	2	US-08-480-784-80 Sequence 80, Appl
25	10.6	62.4	30	2	US-08-483-553-80 Sequence 80, Appl
26	10.6	62.4	30	2	US-08-487-002-80 Sequence 80, Appl
27	10.6	62.4	30	2	US-08-483-554B-80 Sequence 80, Appl

28	10.6	62.4	30	2	US-08-488-011B-80 Sequence 80, Appl
29	10.6	62.4	30	6	PCT-US95-10202-80 Sequence 80, Appl
30	10.6	62.4	30	6	PCT-US95-10203-80 Sequence 80, Appl
31	10.6	62.4	30	6	PCT-US95-10220-80 Sequence 12, Appl
32	10.6	62.4	36	1	US-07-747-785-12 Sequence 6, Appl
33	10.6	62.4	42	1	US-07-885-689A-6 Sequence 15, Appl
34	10.6	62.4	45	4	US-08-495-695B-15 Sequence 15, Appl
35	10.6	62.4	45	6	PCT-US94-14436-15 Sequence 3, Appl
36	10.6	62.4	56	1	US-07-854-845B-3 Sequence 13, Appl
37	10.6	62.4	75	1	US-07-766-351-13 Sequence 13, Appl
38	10.6	62.4	75	1	US-08-059-032-13 Sequence 13, Appl
39	10.6	62.4	75	6	PCT-US91-07290-13 Sequence 13, Appl
40	10.6	62.4	87	2	US-08-474-140-13 Sequence 13, Appl
41	10.6	62.4	87	2	US-08-477-630-13 Sequence 13, Appl
42	10.6	62.4	87	2	US-08-472-293-13 Sequence 13, Appl
43	10.6	62.4	87	2	US-08-474-545-13 Sequence 13, Appl
44	10.6	62.4	87	2	US-08-478-341-13 Sequence 13, Appl
45	10.6	62.4	90	1	US-08-123-702-19 Sequence 19, Appl

#### ALIGNMENTS

RESULT 1  
US-08-479-783A-10  
Sequence 10, Application US/08479783A  
Patent No. 5668264  
GENERAL INFORMATION:  
APPLICANT: NEBOJSA JANJIC  
APPLICANT: LARRY GOLD  
TITLE OF INVENTION: HIGH AFFINITY PDGF NUCLEIC  
TITLE OF INVENTION: ACID LIGANDS  
NUMBER OF SEQUENCES: 90  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Swanson and Bratschun, L.L.C.  
STREET: 8400 East Prentice Avenue, Suite #200  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.4 Mb storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/479,783A  
FILING DATE: 7-JUNE-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/714,131  
FILING DATE: 10-JUNE-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/931,473  
FILING DATE: 17-AUGUST-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/964,624  
FILING DATE: 21-OCTOBER-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/117,991  
FILING DATE: 8-SEPTEMBER-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/536,428  
FILING DATE: 11-JUNE-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Diane H. McLearn  
REGISTRATION NUMBER: 33,960  
REFERENCE/DOCKET NUMBER: NEX42-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 793-3333  
TELEFAX: (303) 793-3433  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:

LENGTH: 86 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-479-783A-10

Query Match 74.1%; Score 12.6; DB 1; Length 86;  
Best Local Similarity 70.6%; Pred. No. 1.1e+02;  
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tgytngcrtcnacytg 17  
||| ||: || ||: ||  
Db 50 TGGTAGCGTCACCTTG 66

## RESULT 2

US-08-479-725-10  
; Sequence 10, Application US/08479725  
; Patent No. 5674685  
; GENERAL INFORMATION:  
; APPLICANT: NEBOJSA JANJIC  
; APPLICANT: LARRY GOLD  
; TITLE OF INVENTION: HIGH AFFINITY PDGF NUCLEIC  
; TITLE OF INVENTION: ACID LIGANDS  
; NUMBER OF SEQUENCES: 90  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Swanson and Bratschun, L.L.C.  
; STREET: 8400 East Prentice Avenue, Suite #200  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.4 Mb storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/479,725  
; FILING DATE: 7-JUNE-1995  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/714,131  
; FILING DATE: 10-JUNE-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/931,473  
; FILING DATE: 17-AUGUST-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/964,624  
; FILING DATE: 21-OCTOBER-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/117,991  
; FILING DATE: 8-SEPTEMBER-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/536,428  
; FILING DATE: 11-JUNE-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Diane H. McClearn  
; REGISTRATION NUMBER: 33,960  
; REFERENCE/DOCKET NUMBER: NEX42-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 793-3333  
; TELEFAX: (303) 793-3433  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 86 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
US-08-479-725-10

Query Match 74.1%; Score 12.6; DB 1; Length 86;  
Best Local Similarity 70.6%; Pred. No. 1.1e+02;  
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tgytngcrtcnacytg 17  
||| ||: || ||: ||  
Db 50 TGGTAGCGTCACCTTG 66

## RESULT 3

US-08-618-693-10  
; Sequence 10, Application US/08618693  
; Patent No. 5723594  
; GENERAL INFORMATION:  
; APPLICANT: NEBOJSA JANJIC  
; APPLICANT: LARRY GOLD  
; TITLE OF INVENTION: HIGH AFFINITY PDGF NUCLEIC  
; TITLE OF INVENTION: ACID LIGANDS  
; NUMBER OF SEQUENCES: 96  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Swanson and Bratschun, L.L.C.  
; STREET: 8400 East Prentice Avenue, Suite #200  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.4 Mb storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect 6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/618,693  
; FILING DATE: 20 MARCH 1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/479,783  
; FILING DATE: 7-JUNE-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Barry J. Swanson  
; REGISTRATION NUMBER: 33,215  
; REFERENCE/DOCKET NUMBER: NEX42/CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 793-3333  
; TELEFAX: (303) 793-3433  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 86 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
US-08-618-693-10

Query Match 74.1%; Score 12.6; DB 2; Length 86;  
Best Local Similarity 70.6%; Pred. No. 1.1e+02;  
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tgytngcrtcnacytg 17  
||| ||: || ||: ||  
Db 50 TGGTAGCGTCACCTTG 66

RESULT 4  
US-08-447-169A-168  
; Sequence 168, Application US/08447169A  
; Patent No. 5811533  
; GENERAL INFORMATION:  
; APPLICANT: JANJIC, N. and GOLD, L.  
; TITLE OF INVENTION: HIGH-AFFINITY OLIGONUCLEOTIDE





ADDRESSEE: Fish & Richardson, P.C.  
STREET: 2200 Sand Hill Road, Suite 100  
CITY: Menlo Park  
STATE: CA  
COUNTRY: US  
ZIP: 94025  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/859,998  
FILING DATE: 21-MAY-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Field, Bret E.  
REGISTRATION NUMBER: 37,620  
REFERENCE/DOCKET NUMBER: 09096/002001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-322-5070  
TELEFAX: 415-854-0875  
INFORMATION FOR SEQ ID NO: 1003:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
FEATURE:  
OTHER INFORMATION: oligonucleotide primer  
US-08-859-998-1003

Query Match 65.9%; Score 11.2; DB 4; Length 27;  
Best Local Similarity 62.5%; Pred. No. 6.3e+02;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 tgyttgcttcnacyt 16  
||:| | | | | | | | | |  
DB 17 TGCTTTCATCCACCT 2

RESULT 7  
US-08-207-901-50  
Sequence 50, Application US/08207901  
Patent No. 5629153  
GENERAL INFORMATION:  
APPLICANT: Urdca, Michael S.  
TIME OF INVENTION: use of DNA-DEPENDENT RNA POLYMERASE  
TITLE OF INVENTION: TRANSCRIPTS AS REPORTER MOLECULES FOR SIGNAL  
NUMBER OF INVENTION: AMPLIFICATION IN NUCLEIC ACID HYBRIDIZATION ASSAYS  
NUMBER OF SEQUENCES: 120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CHIRON CORPORATION  
STREET: 4560 Horton Street  
CITY: Emeryville  
STATE: California  
COUNTRY: USA  
ZIP: 94608  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/207,901  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/639,560B  
FILING DATE: 10-JAN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldman, Kenneth M.  
REGISTRATION NUMBER: 34,174  
REFERENCE/DOCKET NUMBER: 00081.002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2719  
TELEFAX: 510-655-3542  
TELEX: N/A  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-207-901-50

Query Match 65.9%; Score 11.2; DB 1; Length 50;  
Best Local Similarity 62.5%; Pred. No. 6.5e+02;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 tgyttgcttcnacyt 16  
||:| | | | | | | | | |  
DB 21 TGTTTAGCATCTCTT 36

RESULT 8  
US-08-454-557C-74/C  
Sequence 74, Application US/08454557C  
Patent No. 5830670  
GENERAL INFORMATION:  
APPLICANT: de la Monte, Suzanne  
TIME OF INVENTION: Neutral Thread Protein Gene Expression and Detection  
TITLE OF INVENTION: of Alzheimer's Disease  
NUMBER OF SEQUENCES: 121  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Avenue, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/454,557C  
FILING DATE: 30-MAY-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Ludwig, Steven R.  
REGISTRATION NUMBER: 36,203  
REFERENCE/DOCKET NUMBER: 0609.3840003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 74:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 53 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
US-08-454-557C-74

Query Match 65.9%; Score 11.2; DB 3; Length 53;  
Best Local Similarity 62.5%; Pred. No. 6.5e+02;

Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 gytngcrtcnactg 17  
1:1 11:11 11:11  
Db 32 GCTATGATCACTTG 17

## RESULT 9

US-08-340-426D-74/c

; Sequence 74, Application US/08340426D  
; Patent No. 5948634

; GENERAL INFORMATION:

; APPLICANT: de la Monte, Suzanne

; APPLICANT: Mands, Jack R.

; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection

; NUMBER OF SEQUENCES: 121

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein &amp; Fox P.L.L.C.

; STREET: 1100 New York Avenue, Suite 600

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005-3934

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/340,426D

; FILING DATE: 14-NOV-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Ludwig, Steven R.

; REGISTRATION NUMBER: 36,203

; REFERENCE/DOCKET NUMBER: 0609.3840002

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 74:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 53 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: both

; TOPOLOGY: both

US-08-340-426D-74

Query Match 65.9%; Score 11.2; DB 4; Length 53;  
Best Local Similarity 62.5%; Pred. No. 6.5e+02;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 gytngcrtcnactg 17  
1:1 11:11 11:11  
Db 32 GCTATGATCACTTG 17

## RESULT 10

US-08-450-673C-74/c

; Sequence 74, Application US/08450673C  
; Patent No. 5948888

; GENERAL INFORMATION:

; APPLICANT: de la Monte, Suzanne

; APPLICANT: Mands, Jack R.

; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection

; NUMBER OF SEQUENCES: 121

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein &amp; Fox P.L.L.C.

; STREET: 1100 New York Avenue, Suite 600

; CITY: Washington

; STATE: D.C.

COUNTRY: U.S.A.  
ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/450,673C

FILING DATE: 30-MAY-1995

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Ludwig, Steven R.

REGISTRATION NUMBER: 36,203

REFERENCE/DOCKET NUMBER: 0609.3840004

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 74:

SEQUENCE CHARACTERISTICS:

LENGTH: 53 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: both

US-08-450-673C-74

Query Match 65.9%; Score 11.2; DB 4; Length 53;  
Best Local Similarity 62.5%; Pred. No. 6.5e+02;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 gytngcrtcnactg 17  
1:1 11:11 11:11  
Db 32 GCTATGATCACTTG 17

## RESULT 11

PCT-US95-17111A-74/c

; Sequence 74, Application PC/TUS9517111A

; GENERAL INFORMATION:

; APPLICANT: de la Monte, Suzanne

; APPLICANT: Mands, Jack R.

; TITLE OF INVENTION: Neural Thread Protein Gene Expression and

; NUMBER OF SEQUENCES: 121

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein &amp; Fox P.L.L.C.

; STREET: 1100 New York Avenue, Suite 600

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/17111A

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/340,426

FILING DATE: 14-NOV-1994

ATTORNEY/AGENT INFORMATION:

NAME: Ludwig, Steven R.

REGISTRATION NUMBER: 36,203

REFERENCE/DOCKET NUMBER: 0609.3840002

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 74:

SEQUENCE CHARACTERISTICS:

LENGTH: 53 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
PCT-US95-17111A-74

Query Match 65.9%; Score 11.2; DB 6; Length 53;  
Best Local Similarity 62.5%; Pred. No. 6.5e+02;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 2 gytgtcrtcnacgt 17  
1:1 11:11 11:11  
Db 32 GGTATCATCATCACTG 17

RESULT 12  
US-08-564-955-60  
Sequence 60, Application US/08564955  
Patent No. 5811238  
GENERAL INFORMATION:  
APPLICANT: STEMMER, WILLEM P.C.  
APPLICANT: CRAMER, ANDREAS M.  
TITLE OF INVENTION: METHODS FOR GENERATING POLYNUCLEOTIDES  
TITLE OF INVENTION: HAVING DESIRED CHARACTERISTICS BY ITERATIVE SELECTION AND  
NUMBER OF SEQUENCES: 67  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW  
STREET: TWO EMBARCADERO CENTER, 8TH FLOOR  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: U.S.A.  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/564,955  
FILING DATE: 30-NOV-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/198,431  
FILING DATE: 17-FEB-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/537,874  
FILING DATE: 30-OCT-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/02126  
FILING DATE: 17-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: DUNN, TRACY J.  
REGISTRATION NUMBER: 34,587  
REFERENCE/DOCKET NUMBER: 16528J-014611US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 60:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-564-955-60

Query Match 65.9%; Score 11.2; DB 2; Length 54;  
Best Local Similarity 62.5%; Pred. No. 6.5e+02;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 1 tgytngcrtcnacgt 16

Db 7 TCCTGCGCATCACT 22  
11:1 11:11 11:1

RESULT 13  
US-08-564-955-61  
Sequence 61, Application US/08564955  
Patent No. 5811238  
GENERAL INFORMATION:  
APPLICANT: STEMMER, WILLEM P.C.  
APPLICANT: CRAMER, ANDREAS M.  
TITLE OF INVENTION: METHODS FOR GENERATING POLYNUCLEOTIDES  
TITLE OF INVENTION: HAVING DESIRED CHARACTERISTICS BY ITERATIVE SELECTION AND  
NUMBER OF SEQUENCES: 67  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW  
STREET: TWO EMBARCADERO CENTER, 8TH FLOOR  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: U.S.A.  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/564,955  
FILING DATE: 30-NOV-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/198,431  
FILING DATE: 17-FEB-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/537,874  
FILING DATE: 30-OCT-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/02126  
FILING DATE: 17-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: DUNN, TRACY J.  
REGISTRATION NUMBER: 34,587  
REFERENCE/DOCKET NUMBER: 16528J-014611US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 61:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 57 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-564-955-61

Query Match 65.9%; Score 11.2; DB 2; Length 57;  
Best Local Similarity 62.5%; Pred. No. 6.6e+02;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 1 tgytngcrtcnacgt 16  
11:1 11:11 11:1  
Db 7 TCCTGCGCATCACT 22

RESULT 14  
US-08-537-874-58  
Sequence 58, Application US/08537874  
Patent No. 5830721  
GENERAL INFORMATION:  
APPLICANT: STEMMER, WILLEM P.C.  
APPLICANT: CRAMER, ANDREAS  
TITLE OF INVENTION: DNA Mutagenesis by Random Fragmentation

```
; TITLE OF INVENTION: and Reassembly
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/537,874
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP PCT/US95/02126
; FILING DATE: 17-FEB-1995
; APPLICATION NUMBER: US 08/198,431
; FILING DATE: 17-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Liebeschuetz, Joe
; REGISTRATION NUMBER: 37,505
; REFERENCE/DOCKET NUMBER: 018097-014610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 57 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-537-874-58

Query Match          65.9%; Score 11.2; DB 3; Length 57;
Best Local Similarity 62.5%; Pred. No. 6.6e+02;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 tgyttingcrtcnacyt 16
   ||:| ||:| ||:|
Db 7 TGCTGCGCATCCACCT 22

RESULT 15
US-08-537-874-59
; Sequence 59, Application US/08537874
; Patent No. 5830721
; GENERAL INFORMATION:
; APPLICANT: Stemmer, Willem P.C.
; TITLE OF INVENTION: DNA Mutagenesis by Random Fragmentation
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/537,874
; FILING DATE:
```

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP PCT/US95/02126
; FILING DATE: 17-FEB-1995
; APPLICATION NUMBER: US 08/198,431
; FILING DATE: 17-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Liebeschuetz, Joe
; REGISTRATION NUMBER: 37,505
; REFERENCE/DOCKET NUMBER: 018097-014610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 57 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-537-874-59

Query Match          65.9%; Score 11.2; DB 3; Length 57;
Best Local Similarity 62.5%; Pred. No. 6.6e+02;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 tgyttingcrtcnacyt 16
   ||:| ||:| ||:|
Db 7 TGCTGCGCATCCACCT 22
```

Search completed: June 4, 2000, 16:09:39  
Job time: 28058 sec

\_\_\_\_\_

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:05:26 ; Search time 1236.38 Seconds  
(without alignments)  
-15.736 Million cell updates/sec

Title: US-09-164-714-4  
Perfect score: 20  
Sequence: 1 gcnagayncarcnctaygt 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 segs, -486395729 residues

Total number of hits satisfying chosen parameters: 356616

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_em:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pl4:\*  
11: gb\_pl5:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vl:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_or:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl1:\*  
27: em\_pl2:\*  
28: em\_ro:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vl:\*  
32: gb\_htg1:\*  
33: gb\_htg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pl4:\*  
41: gb\_htg3:\*  
42: gb\_htg4:\*  
43: gb\_htg5:\*  
44: gb\_htg6:\*

45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pl5:\*  
52: gb\_pl6:\*  
53: gb\_htg9:\*  
54: gb\_htg10:\*  
55: gb\_htg11:\*  
56: gb\_htg12:\*  
57: gb\_htg13:\*  
58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	12.2	61.0	32	5	A69250	A69250 Sequence 8
2	12.2	61.0	32	5	A71927	A71927 Sequence 8
3	11.6	58.0	30	5	A17053	A17053 Oligonucleo
4	11.6	58.0	30	5	A17453	A17453 Oligonucleo
5	11.6	58.0	30	5	AR014428	AR014428 Sequence
6	11.6	58.0	30	5	111871	111871 Sequence 37
7	11.6	58.0	71	5	AR017688	AR017688 Sequence
8	11.4	57.0	21	5	AR030742	AR030742 Sequence
9	11.4	57.0	21	5	AR030743	AR030743 Sequence
10	11.4	57.0	39	5	AR022077	AR022077 Sequence
11	11.4	57.0	39	5	192867	192867 Sequence 41
12	11.4	57.0	60	5	A19037	A19037 Nucleotide
13	11.4	57.0	60	5	A19038	A19038 Nucleotide
14	11.4	57.0	75	5	125131	125131 Sequence 21
15	11.4	57.0	75	11	HS091211	U91211 Homo sapien
16	11.4	57.0	90	5	125143	125143 Sequence 39
17	11.2	56.0	30	34	DRORRN	K01289 D.melanogas
18	11.2	56.0	52	14	SYNANVA4	M60112 Avian neovl
19	11.2	56.0	76	16	AB003918	AB003918 HepaticLis
20	11.2	56.0	85	5	A79759	A79759 Sequence 7
21	11.2	56.0	85	5	A79760	A79760 Sequence 8
22	11.2	56.0	90	12	RATVPN	K02434 Rat vasopre
23	11.2	56.0	90	12	RATVPNP	K01495 Rat prepro
24	11.2	56.0	97	5	114127	114127 Sequence 41
25	11.2	55.0	25	5	111740	111740 Sequence 11
26	11.2	55.0	27	5	AR039342	AR039342 Sequence
27	11.2	55.0	31	5	E02911	E02911 Oligonucleo
28	11.2	55.0	32	5	AR003523	AR003523 Sequence
29	11.2	55.0	32	5	162417	162417 Sequence 23
30	11.2	55.0	34	5	186938	186938 Sequence 27
31	11.2	55.0	35	5	119619	119619 Sequence 8
32	11.2	55.0	40	5	E06658	E06658 DNA probe t
33	11.2	55.0	43	5	100832	100832 Sequence 3
34	11.2	55.0	43	5	100841	100841 Sequence 4
35	11.2	55.0	43	5	106027	106027 Sequence 1
36	11.2	55.0	46	5	117466	117466 Sequence 5
37	11.2	55.0	49	12	MUSTGHP	M22386 Mouse 19 ge
38	11.2	55.0	53	5	AR062116	AR062116 Sequence
39	11.2	55.0	56	5	A48236	A48236 Sequence 15
40	11.2	55.0	72	14	AF167289	AF167289 Cloning v
41	11.2	55.0	79	5	AR042663	AR042663 Sequence
42	11.2	55.0	79	5	AR064796	AR064796 Sequence
43	11.2	55.0	84	11	HS243870	A0245870 Homo Sapl
44	11.2	55.0	94	9	H0MFAH11	L14667 Human tumor
45	10.8	54.0	57	5	A18403	A18403 signal sequ

## ALIGNMENTS

RESULT	1				
LOCUS	A69250				
DEFINITION	A69250	32 bp	DNA		
ACCESSION	A69250	Sequence 8 from Patent WO9803644.			
VERSION	A69250.1	GI:4760109			
KEYWORDS	.				
SOURCE	unidentified.				
ORGANISM	unidentified				
REFERENCE	unclassified.				
AUTHORS	1 (bases 1 to 32)				
TITLE	Sommer, B. and Staufenbiel, M.				
JOURNAL	TRANSGENIC ANIMAL MODEL FOR ALZHEIMER DISEASE				
	Patent: WO 9803644-A 29-JAN-1998;				
	CIBA GEIGY AG (CH)				
FEATURES	Location/Qualifiers				
source	1..32				
	/organism="unidentified"				
	/db_xref="taxon:32644"				
BASE COUNT	12 a	11 c	3 g	6 t	
ORIGIN					

	Query Match	Score 11.2	DB 5	length 32
	Best Local Similarity	61.1%	Pred. 1.1e+04	
	Matches 11, Conservative	3	Mismatches 4	Indels 0
				Gaps 0
Q7	1 gcnagcgcncarcnctay	18		
	1111111111111111			
Db	7 GCAGATGTCACACCAATAC	24		

	RESULT	2				
	A71927					
	LOCUS					
	DEFINITION	A71927	32 bp	DNA		11-MAY-1999
	ACCESSION	A71927	Sequence 8 from Patent WO960803200.			
	VERSION	A71927.1	GI:4807985			
	KEYWORDS	.				
SOURCE		unidentified.				
ORGANISM		unidentified				
REFERENCE		unclassified.				
AUTHORS		1 (bases 1 to 32)				
TITLE		Baudu P., Riviere M., Audonnet J. and Bouchardon A. POLYNUCLEOTIDE VACCINE FORMULA, PARTICULARLY FOR TREATING BOVINE RESPIRATORY DISEASE				
JOURNAL		Patent: WO 9803200-A 29-JAN-1998; BAUDU PHILIPPE (FR) Other publication FR 2751229 19980123.				
COMMENT FEATURES						
source		location/Qualifiers				
		. . 32				
		/organism=unidentified"				
		/db_xref=taxon:32644"				
BASE COUNT		12 a	11 c	3 g	6 t	
ORIGIN						

	Query Match	Best Local Similarity	Matches	Score 12.2:	DB 5:	length 32:
		61.1%		Pred. No. 1,1e+04:		
			Conservative 11:	Mismatches 3:	Indels 4:	Gaps 0:
Oy	1 gcagagcncarcntay	18				
	:	:				
db	7 GCAGATGTCACACCATAC	24				

RESULT	3			
AL7053				
LOCUS	AL7053	30 bp	DNA	23-MAR-1994
DEFINITION	oligonucleotide seq	ID No: 37.		
ACCESSION	AL7053			
VERSION	AL7053.1	GI:512847		

KEYWORDS	. synthetic construct.			
SOURCE	synthetic construct			
ORGANISM	artificial sequence.			
REFERENCE	1 (bases 1 to 30)			
AUTHORS	Camble,R., Wilkinson,A.J., Carr,H. and Timms,D.			
TITLE	Polypeptides			
JOURNAL	Patent; EP 0459630-A 38 04-DEC-1991;			
	IMPERIAL CHEMICAL INDUSTRIES PLC; ZENECA LIMITED			
FEATURES	Location/Qualifiers			
source	1..30			
	/organism="synthetic construct"			
	/db_xref="taxon:32630"			
BASE COUNT	5	a	11	c
ORIGIN			8	g
				6
				t

Query Match	58.0%	Score 11.6:	DB 5	Length 30:
Best Local Similarity	64.7%	Pred. 2.5e+04:		
Matches	11:	conservative	2:	mismatches 4:
				Indels 0:
				Gaps 0:
Qy	1	gcagagcncarcncta	17	
	1	111111111111	11	
db	12	gcagagcgcacacgccta	28	

RESULT	4			
A17453				
LOCUS	A17453	30 bp	DNA	PAT
DEFINITION	Oligonucleotide seq	ID No:37.		18-MAR-1994
ACCESSION	A17453			
VERSION	A17453.1	GI:490004		
KEYWORDS	.			
SOURCE	synthetic construct.			
ORGANISM	artificial sequence.			
REFERENCE	1 (bases 1 to 30)			
AUTHORS	Camble,R., Timms,D. and Wilkinson,A.J.			
TITLE	Continuous release pharmaceutical compositions comprising a polypeptide covalently conjugated to a water soluble polymer			
JOURNAL	Patent: EP 0473268-A 39 04-MAR-1992.			
FEATURES	IMPERIAL CHEMICAL INDUSTRIES PLC; ZENECA LIMITED			
source	Location/Qualifiers			
	1..30			
	/organism="Synthetic construct"			
	/db_xref="taxon:37630"			
BASE COUNT	5 a 11 c 8 g 6 t			
ORIGIN				

Query Match	58.0%	Score 11.6	DB 5	Length 30
Best Local Similarity	64.7%	Pred. No. 2.5e+04		
Matches 11; Conservative	2; Mismatches 4; Indels 0; Gaps 0;			
QY	1	gcagagcncarcnta	17	
	11	111111111111	11	
Db	12	GCAGAGCGCCACGGCTA	28	

RESULT	5				
AR014428					
LOCUS	AR014428	30 bp	DNA	PAT	05-DEC-1998
DEFINITION	Sequence	37	from patent	US 5773581.	
ACCESSION	AR014428				
VERSION	AR014428.1	GI:3971882			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 30)				
TITLE	Camble, R., Timms, D. and Wilkinson, A. James.				
	Conjugate of a solution stable G-Csf derivative and a water-soluble polymer				



JOURNAL Patent: US 5773581-A 37 30-JUN-1998;  
 FEATURES Location/Qualifiers  
 source 1..30 /Organism="unknown"  
 BASE COUNT 5 a 11 c 8 g 6 t  
 ORIGIN

Query Match 58.0%; Score 11.6; DB 5; Length 30;  
 Best Local Similarity 64.7%; Pred. No. 2.5e+04;  
 Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 gngaygncarcnta 17  
 |||||  
 Db 12 GCAGACGCGCGCTA 28

RESULT 6  
 LOCUS 11871 30 bp DNA PAT 26-JUL-1995  
 DEFINITION Sequence 37 from Patent US 5416195.  
 ACCESSION 11871  
 VERSION 11871.1 GI:909314  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 30)  
 AUTHORS Cambie,R., Carr,H., Timms,D. and Wilkinson,A.J.  
 TITLE Polypeptide derivatives of granulocyte colony stimulating factor  
 JOURNAL Patent: US 5416195-A 37 16-MAY-1995;  
 FEATURES Location/Qualifiers  
 source 1..30 /Organism="unknown"

BASE COUNT 5 a 11 c 8 g 6 t  
 ORIGIN

Query Match 58.0%; Score 11.6; DB 5; Length 30;  
 Best Local Similarity 64.7%; Pred. No. 2.5e+04;  
 Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 gngaygncarcnta 17  
 |||||  
 Db 12 GCAGACGCGCGCTA 28

RESULT 7  
 LOCUS AR017688 71 bp DNA PAT 05-DEC-1998  
 DEFINITION Sequence 89 from patent US 5780228.  
 ACCESSION AR017688  
 VERSION AR017688.1 GI:3973291  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 71)  
 AUTHORS Parma,D.H., Hicke,B., Bridonneau,P. and Gold,L.  
 TITLE High affinity nucleic acid ligands to lectins  
 JOURNAL Patent: US 5780228-A 89 14-JUL-1998;  
 FEATURES Location/Qualifiers  
 source 1..71 /Organism="unknown"

BASE COUNT 19 a 15 c 26 g 11 t  
 ORIGIN

Query Match 58.0%; Score 11.6; DB 5; Length 71;  
 Best Local Similarity 57.9%; Pred. No. 2.6e+04;  
 Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 cngaygncarcntayt 20

Db 36 CAGACGCGCTACTTATGT 18

RESULT 8  
 LOCUS AR030742/c 21 bp DNA PAT 29-SEP-1999  
 DEFINITION Sequence 28 from patent US 5861298.  
 ACCESSION AR030742  
 VERSION AR030742.1 GI:5943956  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 21)  
 AUTHORS Adams,M.D., Blake,J.A., Debouck,C.M., Drake,F.H., Fitzgerald,L.M.,  
 Fraser,C.M., Gowen,M., Hastings,G.A., Kirkness,E.F., Lee,N.H. and  
 Rood,J.  
 TITLE Cathepsin K gene  
 JOURNAL Patent: US 5861298-A 28 19-JAN-1999;  
 FEATURES Location/Qualifiers  
 source 1..21 /Organism="unknown"

BASE COUNT 5 a 5 c 5 g 6 t  
 ORIGIN

Query Match 57.0%; Score 11.4; DB 5; Length 21;  
 Best Local Similarity 60.0%; Pred. No. 3.3e+04;  
 Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 gngaygncarcntayt 20  
 |||||  
 Db 21 GAAGATGCTTACCATATGT 2

RESULT 9  
 LOCUS AR030743 21 bp DNA PAT 29-SEP-1999  
 DEFINITION Sequence 29 from patent US 5861298.  
 ACCESSION AR030743  
 VERSION AR030743.1 GI:5943957  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 21)  
 AUTHORS Adams,M.D., Blake,J.A., Debouck,C.M., Drake,F.H., Fitzgerald,L.M.,  
 Fraser,C.M., Gowen,M., Hastings,G.A., Kirkness,E.F., Lee,N.H. and  
 Rood,J.  
 TITLE Cathepsin K gene  
 JOURNAL Patent: US 5861298-A 29 19-JAN-1999;  
 FEATURES Location/Qualifiers  
 source 1..21 /Organism="unknown"

BASE COUNT 6 a 5 c 5 g 5 t  
 ORIGIN

Query Match 57.0%; Score 11.4; DB 5; Length 21;  
 Best Local Similarity 60.0%; Pred. No. 3.3e+04;  
 Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 gngaygncarcntayt 20  
 |||||  
 Db 1 GAAGATGCTTACCATATGT 20

RESULT 10  
 LOCUS AR022077/c 39 bp DNA PAT 05-DEC-1998  
 DEFINITION Sequence 41 from patent US 5792456.  
 ACCESSION AR022077

VERSION	AR022077.1	GI:3976139
KEYWORDS		
SOURCE	Unknown.	
ORGANISM	Unknown.	
REFERENCE	Unclassified.	
AUTHORS	1 (bases 1 to 39)	
TITLE	Yelton,D., Glaser,S., Huse,W. and Rosok,M.Joanne.	
JOURNAL	Mutant BR96 antibodies reactive with human carcinomas	
FEATURES	Patent: US 5792456-A 41 11-NOV-1998;	
source	Location/Qualifiers	
	1..39	
BASE COUNT	/organism="unknown"	
ORIGIN	10 a 5 c 15 g 9 t	
Query Match	57.0%; Score 11.4; DB 5;	Length 39;
Best Local Similarity	60.0%; Pred. No.3.4e+04;	
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0		
OY	1 gcngaygcncarcnctatggt 20	
Db	25 GCAGATTCACCTCCATATGT 6	
RESULT 11		
LOCUS	192867/c	PAT 01-DEC-1998
DEFINITION	Sequence 41 from patent US 5728821.	
ACCESION	192867	
VERSION	192867.1 GI:3973737	
KEYWORDS	.	
SOURCE	Unknown.	
ORGANISM	Unknown.	
REFERENCE	Unclassified.	
AUTHORS	1 (bases 1 to 39)	
TITLE	Yelton,D., Glaser,S., Huse,W. and Rosok,M.Joanne.	
JOURNAL	Mutant BR96 antibodies reactive with human carcinomas	
FEATURES	Patent: US 5728821-A 41 17-MAR-1998;	
source	Location/Qualifiers	
	1..39	
BASE COUNT	/organism="unknown"	
ORIGIN	10 a 5 c 15 g 9 t	
Query Match	57.0%; Score 11.4; DB 5;	Length 39;
Best Local Similarity	60.0%; Pred. No.3.4e+04;	
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0		
OY	1 gcngaygcncarcnctatgt 20	
Db	25 GCAGATTCACCTCCATATGT 6	
RESULT 12		
LOCUS	A19037/c	PAT 25-APR-1994
DEFINITION	Nucleotide sequence 17 from patent number EP0464533.	
ACCESION	A19037	
VERSION	A19037.1 GI:513972	
KEYWORDS	.	
SOURCE	unidentified.	
ORGANISM	unclassified	
REFERENCE	unclassified.	
AUTHORS	1 (bases 1 to 60)	
TITLE	Lautier,L., Oquendo,P., Zettlmeissl,G. and Seed,B.	
JOURNAL	Fusionproteins with parts of immunoglobulins, their production and	
USE	Patent: EP 0464533-A 17 08-JAN-1992;	
BEHRINGWERKE Aktiengesellschaft; THE GENERAL HOSPITAL CORPORATION		
location/Qualifiers		
1..60		

BASE COUNT	10 a	23 c	15 g	12 t	
ORIGIN					
Query Match	57.0%;	Score 11.4;	DB 5;	Length 60;	
Best Local Similarity	60.0%;	Pred. No. 3.4e+04;			
Matches 12;	Conservative	2;	Mismatches 6;	Indels 0;	Gaps 0;
QY	1	gcngaygcncarcntaygt 20			
		: : : : :			
DB	51	GCTGACGCCACACGAGAGGT 32			
RESULT 13					
AL9038	Al9038	60 bp	DNA	PAT	25-APR-1994
LOCUS	Nucleotide sequence 18 from patent number EP0464533.				
DEFINITION	Al9038				
ACCESSION	Al9038.1	GI:512290			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS	1	(bases 1 to 60)			
TITLE	Laifner, L., Oquendo, P., Zeitlmeissl, G. and Seed, B. Fusionproteins with parts of immunoglobulins, their production and use				
JOURNAL	Patent: EP 0464533-A 18 08-JAN-1992; BEHRINGWERKE Aktiengesellschaft; THE GENERAL HOSPITAL CORPORATION				
FEATURES					
source					
BASE COUNT	12 a	15 c	23 g	10 t	
ORIGIN					
Query Match	57.0%;	Score 11.4;	DB 5;	Length 60;	
Best Local Similarity	60.0%;	Pred. No. 3.4e+04;			
Matches 12;	Conservative	2;	Mismatches 6;	Indels 0;	Gaps 0;
QY	1	gcngaygcncarcntaygt 20			
		: : : : :			
DB	10	GCTGACGCCACACGAGAGGT 29			
RESULT 14					
125131/c	125131	75 bp	DNA	PAT	07-OCT-1996
LOCUS	Sequence 21 from patent US 5547871.				
DEFINITION	125131				
ACCESSION	125131.1	GI:1605001			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS	1	(bases 1 to 75)			
TITLE	Black, B.C. and Summers, M.D. Heterologous signal sequences for secretion of insect controlling proteins				
JOURNAL	Patent: US 5547871-A 21 20-AUG-1996;				
FEATURES					
source					
BASE COUNT	10 a	22 c	24 g	19 t	
ORIGIN					
Query Match	57.0%;	Score 11.4;	DB 5;	Length 75;	
Best Local Similarity	60.0%;	Pred. No. 3.4e+04;			
Matches 12;	Conservative	2;	Mismatches 6;	Indels 0;	Gaps 0;

OY 1 gcnagaygncarcntaygt 20  
 |||||  
 DB 39 GATGAGGCCGACGACGT 20

RESULT 15  
 HSU91211 75 bp mRNA PRI 03-JUL-1997  
 LOCUS HSU91211/c Homo sapiens T-cell receptor delta chain (TCRDV3J2) mRNA, partial  
 DEFINITION cds.  
 ACCESSION U91211  
 VERSION U91211.1 GI:2239828  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 75)  
 AUTHORS Holmeier,W., Witthoft,T., Hennemann,A., Winter,H.S. and  
 Kagnoff,M.F.  
 TITLE The TCR-delta repertoire in human intestine undergoes  
 JOURNAL characteristic changes during fetal to adult development  
 MEDLINE J. Immunol. 158 (12), 5632-5641 (1997)  
 REFERENCE 97334214  
 2 (bases 1 to 75)  
 AUTHORS Holmeier,W., Witthoft,T., Hennemann,A., Harland,S.W. and  
 Kagnoff,M.F.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-FEB-1997) Department of Medicine, University of  
 Frankfurt, Theodor-Stern Kai #7, Frankfurt 60590, Germany  
 FEATURES  
 source  
 1. /75  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="D01363"  
 /chromosome="14"  
 /tissue\_type="small intestine"  
 /dev\_stage="one day old"  
 <1. >75  
 /gene="TCRDV3J2"  
 <1. >75  
 /gene="TCRDV3J2"  
 <1. >75  
 /note="rearranged; contains CDR3 domain, 13 amino acids  
 (calculation according to: Rock,E.  
 J. Exp. Med.179:323-328,1994)"  
 /codon\_start=1  
 /product="T-cell receptor delta chain"  
 /protein\_id="AAC51428.1"  
 /db\_xref="GI:2239829"  
 /translation="TEDSATYYCASALTGCGDALTAQLF"

BASE COUNT 15 a 20 c 21 g 19 t  
 ORIGIN

Query Match 57.0%; Score 11.4; DB 11; Length 75;  
 Best Local Similarity 60.0%; Pred. No. 3.4e+04;  
 Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;  
 OY 1 gcnagaygncarcntaygt 20  
 |||||  
 DB 35 GCCGACGACGACGACGTAGT 16

Search completed: June 4, 2000, 16:05:28  
 Job time: 27880 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:23:59 ; Search time 322.35 Seconds  
(without alignments)  
15.523 Million cell updates/sec

Title: US-09-164-714-4

Perfect score: 20  
Sequence: 1 gongaygencarcntaygt 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 ; Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 431286

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12.2	61.0	32	1	V49438 Primer AB028 for B
2	12	60.0	50	1	c-fos position 272
3	11.6	58.0	40	1	Oligonucleotide ge
4	11.6	58.0	40	1	Oligonucleotide ge
5	11.6	58.0	53	1	C. limi endogluca
6	11.6	58.0	71	1	L-selectin family
7	11.4	57.0	21	1	Human cathepsin K
8	11.4	57.0	21	1	Human cathepsin K
9	11.4	57.0	39	1	5'-Primer for BR96
10	11.4	57.0	47	1	PCR primer C3634CC
11	11.4	57.0	63	1	Drosophila melano
12	11.4	57.0	63	1	Codon optimised s1
13	11.4	57.0	75	1	Codon optimised D.
14	11.4	57.0	90	1	Fragment B estera
15	11.4	57.0	90	1	AAT optimised seq
16	11.4	57.0	90	1	Oligomer for a cod
17	11.2	56.0	97	1	Sequence of synthe
18	11	55.0	25	1	CPTA panhandle PCR
19	11	55.0	27	1	Human IL-2 recepto
20	11	55.0	29	1	PCR primer p1 to a
21	11	55.0	32	1	Human cDNA clone #
22	11	55.0	32	1	Lambda 37225-3725
23	11	55.0	35	1	Probe for human ba
24	11	55.0	39	1	PCR primer for PET
25	11	55.0	39	1	PCR primer Sq2258
26	11	55.0	40	1	Mycoplasma plasm/V
27	11	55.0	40	1	PCR primer for hum
28	11	55.0	40	1	Probe for 2.5-dike
29	11	55.0	43	1	N84347 Oligonucleotide pr
30	11	55.0	43	1	N82099 Murine osteoproteg
31	11	55.0	44	1	p70295 p71 VH cDNA primer
32	11	55.0	46	1	Murine osteoproteg
33	11	55.0	46	1	V70301 Hepatoma AS-30D Ty
34	11	55.0	49	1	T80499

35	11	55.0	53	1	063430 Lambda phage repli
36	11	55.0	53	1	V70296 Murine osteoproteg
37	11	55.0	56	1	T16105 Plasmid pBR322 PCR
38	11	55.0	75	1	Q11149 Probe GTR-1 based
39	11	55.0	81	1	X32352 Artificial NS4 mos
40	11	55.0	83	1	063427 Lambda phage repli
41	11	55.0	83	1	X32356 Artificial NS4 mos
42	10.6	53.0	19	1	W72009 Electronic perturb
43	10.6	53.0	21	1	035337 PCR primer HIV3B9.
44	10.6	53.0	23	1	V64049 Human leiomyoma DN
45	10.6	53.0	24	1	075804 Trimer oligonucleo

## ALIGNMENTS

RESULT 1	
V49438	
ID	V49438 standard; DNA; 32 BP.
AC	V49438; 1998 (first entry)
DT	28-OCT-1998
DE	Primer AB028 for BRSV protein G gene.
KW	Multivalent vaccine; pathogen; respiratory disease; digestive disease;
KW	cattle; bovine herpesvirus; BHV; bovine respiratory syncytial virus;
KW	BRSV; bovine viral diarrhoea virus; BVDV; bovine parainfluenza virus;
KW	BPTV; vector; glycoprotein; primer; PCR; amplification; ss.
OS	Synthetic.
OS	Bovine respiratory syncytial virus.
PN	FR2751229-A1.
PD	23-JAN-1998.
PF	19-JUL-1996; 009403.
PR	19-JUL-1996; FR-009403.
PA	(INMR ) RHONE MERIEUX SA.
PI	Audonnet JCF, Baudu P, Bouchardon A, Riviere MEA;
DR	WPI: 98-112829/11.
PT	Multi-valent polynucleotide vaccines against bovine pathogens -
PT	consist of at least 3 plasmids able to express protective antigens
PT	from specified viruses
PS	Example 10; Page 13; 33p; French.
CC	The invention relates to a multivalent vaccine for protecting cattle
CC	against several pathogens, especially pathogens associated with
CC	respiratory and digestive diseases. The pathogens are especially
CC	selected from bovine herpesvirus (BHV), bovine respiratory syncytial
CC	virus (BRSV), bovine viral diarrhoea virus (BVDV) or bovine parainfluenza
CC	virus (BPTV). The vaccines are preferably composed of polynucleotide
CC	sequences encoding 3 antigens, all as part of vectors.
CC	Primers V49438-V49439 were used to PCR amplify the BRSV strain 391-2
CC	protein G gene. The sequence was subcloned into the plasmid pVR1012
CC	to generate plasmid pAB012 for use in the vaccine.
SO	Sequence 32 BP; 12 A; 11 C; 3 G; 6 T;
Query Match	61.0%; Score 12.2; DB 1; Length 32;
Best Local Similarity	61.1%; Pred. No. 3.6e+02;
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;	
QY	1 gongaygencarcntay 18
DB	7 GCAGATGTCACACCATAC 24
RESULT 2	
Q50214	
ID	Q50214 standard; DNA; 50 BP.
AC	Q50214;
DT	06-MAY-1994 (first entry)
DE	c-fos position 2721-2770 INS mutagenic oligonucleotide.
KW	HIV-1; p17gag; inhibition; gag; M1; M2; M3; M4; silent mutation; mRNA;
KW	vector; p17; point mutation; p17M1234; H124 cells; gene replacement;
KW	inhibitory/instability signal; INS; stability; utilisation; vaccine;
KW	interferon; interleukin; fos proto-oncogene protein; growth factor;
KW	env; attenuated; AIDS; AIDS-related disease; latent infection;
KW	gene therapy; Human immunodeficiency virus type 1; p24;

KW long terminal repeat; LTR; ss.  
 OS Synthetic.  
 PN MO9320212-A.  
 PD 14-OCT-1993.  
 PE 29-MAR-1993; U02908.  
 PR 27-MAR-1992; US-858747.  
 PA (USSH ) US SEC DEPT HEALTH.  
 PI Felber BK, Pavlakis GN;  
 DR WPI: 93-336919/42.  
 PT Eliminating inhibitory-instability regions in mRNA to improve  
 PT stability and expression - by making multiple point mutations  
 PT within A-T rich regions  
 PS Example 3; Page 70; 117pp: English.  
 CC The sequences given in Q50202-70 are oligonucleotides which were  
 CC used to illustrate the method of the invention for the mutagenesis of  
 CC inhibitory/instability signals (INS). Mutation in an INS produced  
 CC using oligonucleotides such as these, increases the stability and/or  
 CC utilization of mRNA without changing its protein coding capacity, or  
 CC if the sequence is changed, its function is maintained. Other genes  
 CC encoding such RNA molecules include growth factor, interferon,  
 CC interleukin, fos proto-oncogene protein and HIV env and gag gene  
 CC proteins. Nucleic acid constructs in which INS function has been  
 CC impeded, can be used as vaccines, esp. against AIDS and AIDS-related  
 CC diseases by preventing HIV from establishing a latent infection, as  
 CC is possible using the INS, and thus escaping immune system  
 CC surveillance. The constructs may also be used in gene therapy for  
 CC gene replacement by homologous recombination with a target gene in  
 CC situ. See also q50200-02.  
 SQ Sequence 50 BP; 16 A; 13 C; 13 G; 8 T;

Query Match	60.0%	Score 12;	DB 1;	Length 50;
Best Local Similarity	63.2%	Pred. No. 4.9e+02;		
Matches 12; Conservative	2;	Mismatches 5;	Indels 0;	Gaps 0;
Oy	2	engaygncarcrcntlaygt	20	
db	17	CTGAAGGACAGCCATACGT	35	

RESULT 3  
 Q38054/C  
 ID Q38054 standard; DNA; 40 BP.  
 AC Q38054;  
 DT 07-JUL-1993 (first entry)  
 DE Oligonucleotide gell, for prodn. of synthetic gelonin gene.  
 KW Seed, toxin; plant; cloning; ribosomal; protein synthesis; ss;  
 KW Gelonium multiflorum.  
 OS Synthetic.  
 PN WC9305168-A.  
 PD 18-MAR-1993.  
 PE 21-AUG-1992; U07066.  
 PR 06-SEP-1991; US-735949.  
 PA (REERE-) RES DEV FOUND.  
 PI Beattie KL, Rosenblum MG;  
 WP1: 93-100990/12.  
 PT Synthetic DNA encoding gelonin plant toxin - provides nucleotide  
 FT sequence for synthetic gene for prodn. and cloning  
 PS Example 4; Fig 5; 45pp; English.  
 CC The synthetic gelonin gene based on the sequence of Gelonium  
 CC multiflorum gelonin was prepd. by synthesising a number of  
 CC oligonucleotides corresp. to fragments of the gelonin gene and  
 CC annealing and ligating to assemble the intact gene. The  
 CC oligonucleotides were designed to contain a codon triplet for each  
 CC amino acid in the corresp. gelonin fragment. Gelonin is a ribosomal-  
 CC inactivating plant toxin which inhibits protein synthesis. The  
 CC synthetic form of gelonin provides a plentiful, reproducible source  
 CC of gelonin which may be modified.  
 See also Q38041-82.  
 Q3 Sequence 40 BP; 15 A; 12 C; 9 G; 4 T;

Query Match 58.0%; Score 11.6; DB 1; Length 40;

```

Best Local Similarity 57.9%; Pred. No. 7.9e+02;
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 gcgagycncarcrcntayg 19
    ||||| ||| | :
Db 19 GCCGATGCCACAGCGGTTG 1

```

```

RESULT 4
ID 038072 standard; DNA; 40 BP.
AC 038072;
DT 07-JUL-1993 (first entry)
DE Oligonucleotide gel29, for prodn. of synthetic gelonin gene.
KW Seed; toxin; plant; cloning; ribosomal; protein synthesis; ss;
  Gelonium multiflorum.
OS Synthetic.
PN MO9305168-A.
PD 18-MAR-1993.
PF 21-AUG-1992; U07066.
PR 06-SEP-1991; US-755949.
PI (RERE-) RES DEV FOUND.
PI Beattie RL, Rosenblum MG;
DR WPI: 93-100990/12.
PT Synthetic DNA encoding gelonin plant toxin - provides nucleotide
  sequence for synthetic gene for prodn. and cloning
PS Example 4; Fig 5; 45pp; English.
CC The synthetic gelonin gene based on the sequence of Gelonium
  multiflorum gelonin was prepd. by synthesizing a number of
  oligonucleotides corresp. to fragments of the gelonin gene and
  annealing and ligating to assemble the intact gene. The
  oligonucleotides were designed to contain a codon triplet for each
  amino acid in the corresp. gelonin fragment. Gelonin is a ribosomal-
  inactivating plant toxin which inhibits protein synthesis. The
  synthetic form of gelonin provides a plentiful, reproducible source
  of gelonin which may be modified.
CC See also Q38041-82.
SQ Sequence 40 BP; 7 A; 11 C; 12 G; 10 T;

```

Query Match	58.0%	Score 11.6:	DB 1:	Length 40:
Best Local Similarity	57.9%	Pred. No. 7.9e02:		
Matches 11: Conservative	3:	Mismatches 5:	Indels 0:	Gaps 0:
QY	1 gcngaygcncarcrcntayg 19			
	: : : : : : : : :			
Db	2 GCCGATGCGCAGCGGTGCG 20			

RESULT 5

V29675/c

ID V29675 standard; DNA; 53 BP.

AC V29675;

DT 01-SEP-1998 (first entry)

DE C. fimi endoglucanase A (Cena) gene amplifying primer 15, 3'SAENH.

KW Starch, liquefaction; sweetener; enzyme hybrid; endoglucanase; enzyme

KW cellulose binding domain; CDP; starch processing; alpha-amylase;

KW saccharification; Cena; PCR primer; ss.

OS Synthetic.

OS Cellulomonas fimi.

PN MO9816633-A1.

PD 23-APR-1998.

PR 13-OCT-1997; DK0448.

PR 11-OCT-1996; DK-001130.

PA (NOVO ) NOVO-NORDISK AS.

PI Bisgaard/antzen H, Bjornvad M, Pedersen S, Schuelein M;

DR WPI: 98-251283/22.

PT Liquefaction of starch for, e.g. production of sweeteners -

PT comprises use of enzyme hybrids including cellulose binding domain

PT for starch

PS Example 7; Page 39; 83pp; English.

CC This primer is used for the PCR amplification of the Cellulomonas fimi

CC endoglucanase A (Cena) gene fragment encoding the cellulose binding

CC	domain (CBD). This is used in the construction of enzyme hybrids for
CC	liquefaction of starch. The enzyme hybrids contain amino acid sequences
CC	of alpha-amylase linked to a CBD. The starch is liquefied by treating,
CC	in aqueous medium, with such an enzyme hybrid. A recombinant expression
CC	vector comprising a construct containing isolated DNA encoding enzyme
CC	hybrids with amylolytic activity, promoter and stop signals can be used
CC	to transform host cells for the production of the recombinant enzyme
CC	hybrids. The enzyme hybrids are useful in industrial starch processing
CC	especially for the production of sweeteners. Hybrid enzymes have altered
CC	affinity for substrate and increased activity, resulting in at least 1 of
CC	reduced calcium ion dependence, reduced formation of Maillard reaction
CC	products and reduced effect of alpha-amylase on subsequent
CC	saccharification.
SD	Sequence 53 BP: 11 A: 13 C: 16 G: 13 T:
QY	Query Match 58.0%; Score 11.6; DB 1; Length 53;
Db	Best Local Similarity 57.9%; Pred. No. 8.2e+02;
	Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
	1 gcnagaygcncarcntayg 19
	:    :    :
	51 GCCGACGACACACCCACG 33
RESULT	6
T57762/C	
ID	T57762 standard; RNA: 71 BP.
AC	T57762.
DT	21-NOV-1997 (first entry)
DE	L-selectin family III SELEX ligand F13.57.
KW	Identification: ligand: lectin: SELEX; wheat germ agglutinin: template;
KW	Systematic Evolution of ligands by exponential enrichment; amplification;
KW	primer: PCR; polymerase chain reaction; peritoneal inflammation; ss;
KW	diabetes; lymphocyte trafficking disorder; glomerulonephritis; arthritis.
OS	Synthetic.
FH	Key
FT	Location/Qualifiers
FT	modified_base 1..71
FT	/*tag= a
FT	/mod_base= all C bases are 2' NH2-cytosine
FT	/mod_base= all U bases are 2' NH2-uracil
PN	W09640703-A1.
PD	19-DEC-1996.
PF	05-JUN-1996; U09455.
PR	07-JUN-1995; US-479724.
PR	07-JUN-1995; US-472255.
PR	07-JUN-1995; US-472256.
PR	07-JUN-1995; US-477829.
PA	(NEXS-) NEXSTAR PHARM INC.
PI	Bridonneau P, Gold L, Hicke B, Parma DH;
PI	WPI: 97-077252/07.
PT	Identifying nucleic acid ligands that bind lectin(s) esp.
PT	selectin(s) - by partitioning the ligands from a mixture of nucleic
PT	acids
PS	Claim 35; Page 148; 255pp; English.
CC	The invention relates to the identification of nucleic acid ligands to
CC	a lectin using the Systematic Evolution of ligands by Exponential
CC	enrichment (SELEX) method. The sequences T57740-157790 represent RNA
CC	ligands isolated by the method which bind to L-selectin. The L-selectin
CC	ligands were isolated from a DNA template containing 40 random
CC	nucleotides flanked by fixed 5' and 3' sequences (T58043), which was
CC	amplified using the primers T58044-5. The ligands fall into 13 families
CC	along with a group of unrelated 'orphan' ligands. This ligand binds
CC	L-selectin with a Kd of 75 nM at room temperature. The ligands are
CC	especially useful in the treatment of peritoneal inflammation, diabetes,
CC	lymphocyte trafficking disorders, glomerulonephritis, arthritis, etc.
SD	Sequence 71 BP: 19 A: 15 C: 26 G: 11 U;
QY	Query Match 58.0%; Score 11.6; DB 1; Length 71;
Db	Best Local Similarity 57.9%; Pred. No. 8.4e+02;
	Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

Qy      2  cngaygcnarcntaygt 20
        | | | | | | | | | | | |
Db      36  CAGACGCGCTACTTATGT 18

RESULT 7
V09686/c
ID      V09686 standard; DNA: 21 BP.
AC      V09686;
DE      20-JUL-1998 (first entry)
KW      Human cathepsin K gene exon 5 reverse PCR primer 4R.
KW      Cathepsin K: human; osteoporosis; periodontal disease;
KW      Paget's disease; Gaucher's disease; Alzheimer's disease;
KW      central nervous system inflammation; hyperparathyroidism;
KW      bone degradation; dental implant degradation; metastasis;
KW      diagnosis; therapy; marker; PCR; primer; ss.
OS      Synthetic.
OS      Homo sapiens.
PN      EP-812916-A2.
PD      17-DEC-1997.
PF      19-MAY-1997; 303395.
PR      26-AUG-1996; US-026083.
PR      14-JUN-1996; US-019942.
PR      17-JUN-1996; US-020273.
PR      26-AUG-1996; WO-014026.
PA      (HUMA-) HUMAN GENOME SCI INC.
PA      (GENO-) INST GENOMIC RES.
PA      (SMIK ) SMITHKLINE BEECHAM CORP.
PI      Adams MD, Blake JA, Debouch CM, Drake FH, Fitzgerald LM,
PI      Fraser CM, Gowen M, Hastings GA, Kirness EF, Lee NH,
PI      Rood J;
PI      WPI. 98-034977/04.
PT      DNA encoding human cathepsin K - useful for diagnosing and treating
PT      diseases associated with cathepsin K e.g. osteoporosis, bone
PT      degradation, metastatic tumours, etc
PS      Example 1; Page 50; 84pp; English.
CC      Oligonucleotide 4R comprises a reverse PCR primer from exon 5
CC      (see V09670) of the human cathepsin K gene (see V09660). PCR
CC      primers (see V09679-90) to adjacent exons of the cathepsin K
CC      gene were used in the amplification of human genomic DNA. DNA
CC      sequencing of intron-exon boundaries allowed sequencing of the
CC      cathepsin genomic DNA. DNA encoding human cathepsin K is useful
CC      for the diagnosis and treatment of e.g. osteoporosis, periodontal
CC      disease, Paget's disease, Gaucher's disease, CNS inflammation,
CC      Alzheimer's disease, hyperparathyroidism, bone degradation,
CC      metastatic tumours, and degradation of bone implants and
CC      prostheses, especially dental implants.
SQ      Sequence 21 BP; 5 A; 5 C; 5 G; 6 T;

Query Match      57.0%; Score 11.4; DB 1; Length 21;
Best Local Similarity 60.0%; Pred No. 9.5e+02;
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0.

Qy      1  gcngaygcnarcntaygt 20
        | | | | | | | | | | | |
Db      21  GAAGATGCTACCATATGT 2

RESULT 8
V09687
ID      V09687 standard; DNA: 21 BP.
AC      V09687;
DE      20-JUL-1998 (first entry)
KW      Human cathepsin K gene exon 5, 6 forward PCR primer 5-6F.
KW      Cathepsin K: human; osteoporosis; periodontal disease;
KW      Paget's disease; Gaucher's disease; Alzheimer's disease;
KW      central nervous system inflammation; hyperparathyroidism;
KW      bone degradation; dental implant degradation; metastasis;
KW      diagnosis; therapy; marker; PCR; primer; ss.
OS      Synthetic.
OS      Homo sapiens.
PN      EP-812916-A2.

```







DR MP1: 94.42108/230.  
PT Heterologous signal sequences for secretion of insect controlling  
PS proteins - useful to protect plants from insect pests  
PS Example 2: Page 14, 63pp, English.  
CC The sequences given in Q67700-10 are fragments which were used in the  
CC construction of eight heterologous signal sequence-codon optimised AaIT  
CC toxin gene cassettes. The cassettes were constructed in two pieces, a  
CC "B" fragment unique for each construct, consisting of DNA coding for one  
CC of eight heterologous signal sequences, plus the amino terminal portion  
CC of the toxin coding region, and an "A" fragment which is the same for  
CC each construct and encodes the remainder of the toxin coding gene. Each  
CC of the fragments "A" and "B" were synthesised by annealing a pair of  
CC oligomers containing a 15 bp overlap to complete the double stranded  
CC molecule. Fusion constructs such as these, may be introduced into an  
CC insect virus, such as the baculovirus AcMNPV. The insertion of the AaIT  
CC gene and the heterologous signal sequence into a baculovirus results in  
CC the expression and secretion of the toxin. A susceptible insect which  
CC ingests such a modified baculovirus will cease feeding on plants due to  
CC toxin-induced paralysis at an earlier time than an insect which ingests  
CC a wild-type baculovirus, thus reducing crop damage.  
SQ Sequence 90 bp; 17 A; 25 C; 28 G; 20 T;



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 13:52:59 ; Search time 4521.53 Seconds  
(without alignments)  
17.929 Million cell updates/sec

Title: US-09-164-714-4

Perfect score: 20  
Sequence: 1 gcngaygcncarcntaygt 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 119968

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gd\_est1:\*  
21: gd\_est2:\*  
22: gd\_est3:\*  
23: gd\_est4:\*  
24: gd\_est5:\*  
25: gd\_est6:\*  
26: gd\_est7:\*  
27: gd\_est8:\*  
28: gd\_est9:\*  
29: gd\_est10:\*  
30: gd\_est11:\*  
31: gd\_est12:\*  
32: gd\_est13:\*  
33: gd\_est14:\*  
34: gd\_est15:\*  
35: gd\_est16:\*  
36: gd\_est17:\*  
37: gd\_est18:\*  
38: gd\_est19:\*  
39: gd\_est20:\*  
40: gd\_est21:\*  
41: gd\_est22:\*  
42: gd\_est23:\*  
43: gd\_est24:\*  
44: gd\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result

Query

No.	Score		Match Length	DB	ID	Description
	1	2				
C	1	13.2	66.0	89	42	A1159672
C	2	12.6	63.0	75	36	AA607795
C	3	12.6	63.0	98	35	D63753
C	4	12.2	61.0	70	29	AA177604
C	5	12.2	61.0	79	43	A1191114
C	6	12.2	61.0	96	46	A1439657
C	7	12	60.0	88	35	AA574321
C	8	11.6	58.0	67	42	A1094466
C	9	11.6	58.0	76	40	AA962216
C	10	11.6	58.0	85	63	A1965940
C	11	11.6	58.0	100	39	AA862553
C	12	11.2	56.0	36	21	T61825
C	13	11.2	56.0	73	94	AA025502
C	14	11	55.0	67	39	AA026835
C	15	11	55.0	67	41	A1033279
C	16	11	55.0	82	40	AA983628
C	17	11	55.0	88	35	AA558306
C	18	11	55.0	93	64	AA059703
C	19	11	55.0	93	34	AA471987
C	20	11	55.0	96	43	A1192169
C	21	11	55.0	100	35	AA553188
C	22	11	55.0	100	64	AA066340
C	23	11	55.0	100	83	H08
C	24	11	55.0	100	91	A0847649
C	25	10.6	53.0	46	37	AA719000
C	26	10.6	53.0	46	41	A1035942
C	27	10.6	53.0	49	37	AA717442
C	28	10.6	53.0	67	42	A1089772
C	29	10.6	53.0	76	62	A1879443
C	30	10.6	53.0	82	33	AA426465
C	31	10.6	53.0	82	36	AA600198
C	32	10.6	53.0	85	33	AA445562
C	33	10.6	53.0	85	48	A1625618
C	34	10.6	53.0	84	61	A1827573
C	35	10.6	53.0	95	26	W87137
C	36	10.6	53.0	95	37	AA701524
C	37	10.6	53.0	97	34	AA533448
C	38	10.6	53.0	97	39	AA823068
C	39	10.6	53.0	100	34	AA1234612
C	40	10.6	53.0	100	43	A1188636
C	41	10.6	53.0	100	50	AA033926
C	42	10.4	52.0	40	22	AA55306
C	43	10.4	52.0	49	49	A1334100
C	44	10.4	52.0	52	26	W71856
C	45	10.4	52.0	55	23	R93606

## IGNITIONS

RESULT	1
A1159672/c	
LOCUS	
DEFINITION	A1159672.89 bp mRNA EST 02-OCT-1998
IMAGE:	I033a05.tl Sycargo mouse liver mla Mus musculus cDNA clone IMAGE:I49504 3' similar to gb:X04323_cds1 GAP JUNCTION BETA-1 PROTEIN (HUMAN); gb:M63802 Mus musculus connexin 32 (MOUSE)';, mRNA sequence.
ACCESSION	A1159672
VERSION	A1159672.1 GI:3692854
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 89)
AUTHORS	Marrar,M., Hillier,L., Allen,M., Bowles,N., Dietrich,N., Dubuque,T., Gelsell,S., Kucaba,T., Lacy,M., Le-M., Martin,J., Morris,B., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE	The Washo-HMI Mouse EST Project

Unpublished (1996)  
 On Jan 9, 1998 this sequence version replaced gi:936225.  
 COMMENT

Unpublished (1996)  
 On Jan 9, 1998 this sequence version replaced gi:936225.  
 Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL ; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:9377108

Trace considered overall poor quality  
 Seq primer: custom primer used  
 High quality sequence stop: 1.  
 Location/Qualifiers

FEATURES  
 source

1..89

/organism="Mus musculus"  
 /strain="C57BL"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:149504"  
 /clone\_id="Sugano mouse liver mlia"  
 /sex="female"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Organ: liver; Vector: pME18S-FL3; Site\_1: DralII  
 (CACCATGTG); Site\_2: DralII (CACCATGTG); 1st strand cDNA  
 was primed with an oligo(dt) primer  
 [AAGCGCCCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was  
 ligated to a DralII adaptor (TGTGGCGCTACTGG), digested  
 and cloned into distinct DralII sites of the pME18S-FL3  
 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should  
 be used to isolate the cDNA insert. Size selection was  
 performed to exclude fragments <1.5kb. Library  
 constructed by Dr. Sumio Sugano (University of Tokyo  
 Institute of Medical Science). Custom primers for  
 sequencing: 5' end primer CTCTGCTCTCAAAAGCTGCG and 3' end  
 primer CGACCTGCACTGACGACCA."

BASE COUNT 16 a 24 c 26 g 23 t

ORIGIN

Query Match 66.0%; Score 13.2; DB 42; Length 89;  
 Best Local Similarity 63.2%; Pred. No. 1.2e+03;  
 Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 gengaygencarcantayg 19  
 ||||| ||||| ||||| |||||  
 Db 22 GCCGATGCTCAACCTCATG 4

RESULT 2

AA607795 75 bp mRNA EST 30-SEP-1997  
 LOCUS AA607795  
 DEFINITION v049f02.r1 Barcode mouse irradiated colon MRLRP7 Mus musculus cDNA  
 clone IMAGE:1051243 5', mRNA sequence.  
 ACCESSION AA607795  
 VERSION AA607795.1 GI:2455230  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 75)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Treising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
 The WashU-HMI Mouse EST Project  
 Unpublished (1996)  
 On Apr 14, 1993 this sequence version replaced gi:592809.  
 COMMENT

Unpublished (1996)  
 On Apr 14, 1993 this sequence version replaced gi:592809.  
 Contact: Marra M/Mouse EST Project







KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Homiidae; Homo.  
TITLE 1 (bases 1 to 88)  
COMMENT NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Apr 14, 1993 this sequence version replaced g1:693382.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
DNA Sequencing by: Genome Systems Inc., Greg Lennon, Ph.D.  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL at:  
[www.bio.livnl.gov/bdrip/image/image.html](http://www.bio.livnl.gov/bdrip/image/image.html)  
Insert Length: 1746 Std Error: 0.00  
Seq primer: -40ml3 fwd. RT from Amersham  
High quality sequence stop: 1.  
FEATURES  
Source  
Location/Qualifiers  
1. 88  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:916733"  
/clone\_lib="NCI-CGAP\_P12"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"  
/note="Vector: PAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from  
5,000-10,000 microdissected preneoplastic cells  
histologically-determined to be prostatic intraepithelial  
neoplasia 2 (PIN2) cells. Double-stranded cDNA was  
ligated to EcoRI adaptors, 5 cycles of PCR applied to the  
cDNA with an adaptor-specific primer, and the resulting  
PCR product subcloned into PAMP10 by the UDG-cloning  
method (Life Technologies). Average insert size is 600  
bp. NOTE: Not directionally cloned. This library was  
constructed by David Krizman."

BASE COUNT 21 a 30 c 23 g 14 t  
ORIGIN

Query Match 60.0%; Score 12; DB 35; Length 88;  
Best Local Similarity 63.2%; Pred. No. 5.6e+03;  
Matches 12; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 gcgaagcncarcantaygt 19  
|||:||||| 11:|  
Db 17 GCTGATGCACACGTACG 35

RESULT 8  
AI094466/c 67 bp mRNA EST 10-NOV-1998  
LOCUS ou87f09.s1 Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1 Homo sapiens CDNA clone  
DEFINITION IMAGE:1634825 3' similar to contains MER22.t2 MER22 repetitive  
element; , mRNA sequence.  
ACCESSION AI094466 GI:3433442  
VERSION AI094466.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 67)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Jan 19, 1998 this sequence version replaced g1:2284745.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
This clone is available royalty-free through LINL; contact the  
IMAGE Consortium ([info@image.livnl.gov](mailto:info@image.livnl.gov)) for further information.  
Insert length: 431 Std Error: 0.00  
Seq primer: -40ml3 fwd. RT from Amersham  
High quality sequence stop: 53.  
FEATURES  
Source  
Location/Qualifiers  
1. 67  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1634825"  
/clone\_lib="Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pTR73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from five normalized  
libraries were mixed, and ss circles were made in vitro.  
Following HAP purification, this DNA was used as tracer in  
a subtractive hybridization reaction. The driver was  
PCR-amplified cDNAs from pools of 5,000 clones made from  
the same 5 libraries. The pools consisted of the following  
libraries and clones: Soares NBHSF pool 1:  
309384-310919, 323208-325885 Soares NB2HP pool 1:  
145032-147335, 147720-148103, 148872-149255, 15002 -  
150407, 151176-152327 Soares NB2HP8-9W pool 1:  
758280-760583, 772104-774407 Soares NBHPA pool 1:  
304776-306311, 320136-322823, 326280-326663 Soares NBHOT  
pool 1: 723720-726407, 739080-740999 subtraction by Bento  
Soares and M. Fatima Bernaldo."

BASE COUNT 23 a 13 c 18 g 13 t  
ORIGIN

Query Match 58.0%; Score 11.6; DB 42; Length 67;  
Best Local Similarity 57.9%; Pred. No. 8.6e+03;  
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 cngaagcncarcantaygt 20  
|||:||||| 11:|  
Db 59 CCGTGCACACGTACGT 41

RESULT 9  
AA962216/c 76 bp mRNA EST 15-MAY-1998  
LOCUS om80g03.s1 NCI CGAP Kid3 Homo sapiens CDNA clone IMAGE:1553524 3'  
DEFINITION similar to TR:P97876 P97876 JUN DIMERIZATION PROTEIN 1 JDP-1. , ,  
mRNA sequence.  
ACCESSION AA962216  
VERSION AA962216 GI:3134380  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 76)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Sep 12, 1996 this sequence version replaced g1:1402297.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www.bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality  
 Seq primer: -40m3 fwd, ET from Amersham  
 High quality sequence stop: 1.

## FEATURES

source

location/Qualifiers

```
1..76
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:153524"
/clone_lib="NCI-CGAP_Kid3"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaudo."
M. Fatima Bonaudo. "
15 a 23 c 22 g 16 t
```

## BASE COUNT

ORIGIN

## Query Match

Best Local Similarity 58.0%; Score 11.6; DB 40; Length 76;  
 Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 gccagagcncarcntatg 19  
 |||||:|||||:  
 Db 44 GCTGCTCTCTGCCCTATG 26

## RESULT 10

AI965940

LOCUS sc24b04.y1 Gm-c1013 glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
 Gm-c1013-1544 5' similar to FR:004918 004918 CHLOROPHYLL A-B

DEFINITION BINDING PROTEIN OF LHCI TYPE III PRECURSOR ; mRNA sequence.  
 AI965940

ACCESSION AI965940.1 GI:5760577  
 VERSION EST.

KEYWORDS soybean.  
 SOURCE glycine max

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
 eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
 Glycine.

## REFERENCE

AUTHORS

1 (bases 1 to 85)  
 Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V.,  
 Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J.,  
 Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M.,  
 Bowers, J., Person, B., Swaller, T., Gildons, M., Pape, D., Harvey, N.,  
 Schurr, R., Ritzer, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,  
 McCann, R., Waterston, R. and Wilson, R.  
 Public Soybean EST Project  
 Unpublished (1999)

## TITLE

JOURNAL

COMMENT

On May 18, 1998 this sequence version replaced gi:3138522.

Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project

Washington University School of Medicine  
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800  
 Fax: 314 286 1810

Email: estewatson.wustl.edu  
 This clone is available through: Genome Systems, Inc. 4633 World

Parkway Circle St. Louis, Missouri 63134 For further information  
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
 427-3324 or contact: clones@genomesystems.com or  
 info@genomesystems.com web site: www.genomesystems.com  
 Trace considered overall poor quality  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -40RP from gibco  
 High quality sequence stop: 1.

## FEATURES

source

location/Qualifiers

```
1..85
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1013-1544"
/clone_lib="Gm-c1013"
/tissue_type="Whole seedlings, 2-3 week old seedlings,
greenhouse grown"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
XhoI; This cDNA library was constructed from mRNA isolated
from whole seedlings of 2-3 week old greenhouse grown
plants. The cDNA library was prepared using the Stratagene
pBluescript II XR cDNA library construction kit.
Complementary DNA was synthesized from mRNA using a primer
consisting of a poly (dT) sequence with a XhoI restriction
site. EcoRI adaptors were ligated to the blunt-ended cDNA
fragments followed by XhoI digestion. The cDNA fragments
were directionally cloned into the EcoRI-XhoI restriction
site of the pBluescript vector. The ligated cDNA fragments
were transformed into XL10-Gold host cells. This library
was constructed by Dr. Randy Shoemaker and Dr. John
Erpelting."
20 a 21 c 26 g 18 t
```

## BASE COUNT

ORIGIN

Query Match 58.0%; Score 11.6; DB 63; Length 85;  
 Best Local Similarity 57.9%; Pred. No. 9.3e+03;  
 Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 gccagagcncarcntatg 19  
 |||||:|||||:  
 Db 20 GCTGACCGCGACATATG 38

## RESULT 11

AA862553/C

LOCUS AA862553 100 bp mRNA EST 24-AUG-1998  
 OH44b08.s1 NCI-CGAP\_GC4 Homo sapiens cDNA clone IMAGE:1469463 3'

DEFINITION similar to SW:DDX4\_RAT 064060 DEAD BOX PROTEIN 4 ; mRNA sequence.  
 AA862553

ACCESSION AA862553.1 GI:2955032  
 VERSION EST.

KEYWORDS human.  
 SOURCE Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 100)

## REFERENCE

AUTHORS

Tumor Gene Index  
 Unpublished (1997)

On Sep 12, 1996 this sequence version replaced gi:1402122.  
 Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality  
Insert Length: 1202 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amerisham  
High quality sequence stop: 1.  
Location/Qualifiers

## FEATURES

source

```
1. 100
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1469463"
/clone_lib="NCI_GCAP_GC4"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from 3 pooled
germ cell tumors, and was then primed with a Not I -
oligo(dt) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."
```

BASE COUNT

30 a 18 c 19 g 33 t

ORIGIN

Query Match 58.0%; Score 11.6; DB 39; Length 100;  
Best Local Similarity 57.9%; Pred. No. 9.7e+03;  
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Oy 2 cngaygcncarcntaygt 20

Db 60 CAGATGCTCAACATGATGT 42

## RESULT 12

LOCUS T61825 36 bp mRNA EST 14-FEB-1995  
DEFINITION yb92804.s1 Stragene liver (#937224) Homo sapiens CDNA clone  
IMAGE:78606 3' similar to gb:L11566 60S RIBOSOMAL PROTEIN L18  
(HUMAN); mRNA sequence.

ACCESSION T61825  
VERSION T61825.1 GI:665068  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 36)  
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,  
Chissos,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,  
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,  
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,  
Rohtling,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,  
Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.  
and Marra,M.

Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 693

High quality sequence starts: 1 High quality sequence stops: 1  
Source: IMAGE Consortium, LML. This clone is available royalty-free  
through LML; contact the IMAGE Consortium (info@image.lml.gov)  
for further information. Trace considered overall poor quality  
Insert Length: 693 Std Error: 0.00  
Seq primer: -21ml3  
High quality sequence stop: 1.  
Location/Qualifiers

## source

```
1. 36
/organism="Homo sapiens"
/db_xref="GDB:498351"
/db_xref="taxon:9606"
/clone="IMAGE:78606"
/clone_lib="Stragene liver (#937224)"
/sex="male"
/age="49 years old"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: liver; Vector: pBluescript SK; Site: 1:  
EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. Hepatectomy from normal male caucasian. Average  
insert size: 1.1 kb; Uni-ZAP XR Vector; ~5' adaptor  
sequence: 5' GAATTCGCGACGAC 3' ~3' adaptor sequence: 5'  
CTGCGATTTTCTTTTCTTTTCTTTT 3"
```

BASE COUNT

3 a 9 c 15 g 7 t 2 others

ORIGIN

Query Match 56.0%; Score 11.2; DB 21; Length 36;  
Best Local Similarity 62.5%; Pred. No. 1.2e+04;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 5 aycgcncarcntaygt 20

Db 31 ACGCCAACCCTACGT 16

## RESULT 13

LOCUS A0025502/c 73 bp DNA GSS 14-OCT-1998  
DEFINITION EP(X)1563-5prime Drosophila melanogaster EP line Drosophila  
melanogaster genomic Sequence recovered from 5' end of P element,  
genomic survey sequence.

ACCESSION A0025502  
VERSION A0025502.1 GI:3265854  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster

REFERENCE 1 (bases 1 to 73)  
AUTHORS Rehm,E.J. and Rubin,G.M.  
TITLE The BGP gene disruption project: single EP element insertions  
JOURNAL Unpublished (1998)  
CONTACT: Gerald Rubin

Berkeley Drosophila Genome Project  
University of California, Berkeley  
LSA Building, Berkeley, CA 94720-3200, USA  
Fax: 51064339947  
Email: gerry@fruitfly.berkeley.edu

Sequence recovery method was inverse PCR.  
Sequence orientation is forward strand relative to 5' end of P  
element

The P element insertion position is base 66 in the 73 bases. This  
insertion position refers to the first base of the 8 base target  
recognition sequence.

Class: transposon-tagged.  
Location/Qualifiers  
1. 73  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="Drosophila melanogaster EP line"  
/note="Inverse PCR was performed on Drosophila  
melanogaster strains each of which contains a single EP  
transposable element insertion. (The generation of these  
insertion strains is described in North P, Szabo K, Bailey  
A, Laverly T, Rehm J, Rubin GM, Weigmann K, Milian M, Benes

## FEATURES

Location/Qualifiers



---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:09:39 ; Search time 244.64 Seconds  
(without alignments)  
10.627 Million cell updates/sec

Title: US-09-164-714-4

Perfect score: 20

Sequence: 1 gcagagcncarcntayt 20

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 segs, 64992525 residues

Total number of hits satisfying chosen parameters: 375880

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/2/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/lna/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/lna/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/lna/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/lna/PCTUS.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/lna/Backfillsl.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12	60.0	50	4	US-08-850-049-18
2	12	60.0	50	4	US-08-050-478-18
3	11.8	59.0	68	5	US-08-469-318-99
4	11.8	59.0	68	5	US-08-469-318-100
5	11.8	59.0	68	6	PCT-US95-01185-99
6	11.8	59.0	68	6	PCT-US95-01185-100
7	11.6	58.0	30	1	US-07-734-225A-37
8	11.6	58.0	30	1	US-07-692-995B-37
9	11.6	58.0	30	2	US-08-488-457-37
10	11.6	58.0	33	2	US-08-366-953A-18
11	11.6	58.0	71	2	US-08-472-255A-89
12	11.6	58.0	71	2	US-08-479-724A-89
13	11.6	58.0	71	5	US-08-472-256B-89
14	11.6	58.0	71	6	PCT-US96-09455A-89
15	11.4	57.0	21	3	US-08-852-807-28
16	11.4	57.0	21	3	US-08-852-807-29
17	11.4	57.0	35	4	US-08-370-156-23
18	11.4	57.0	35	2	US-08-285-936-41
19	11.4	57.0	39	2	US-08-487-860-41
20	11.4	57.0	47	6	PCT-US93-11638-11
21	11.4	57.0	63	6	PCT-US94-06079-41
22	11.4	57.0	75	1	US-08-009-265-21
23	11.4	57.0	86	4	US-08-477-527A-94
24	11.4	57.0	86	4	US-08-481-710-94
25	11.4	57.0	86	6	PCT-US96-09537-94
26	11.4	57.0	90	1	US-08-009-265-39
27	11.2	56.0	36	6	PCT-US95-15800-13

c 28	11.2	56.0	36	6	PCT-US95-15800-14	Sequence 14, Appl
c 29	11.2	56.0	97	1	US-07-750-080A-41	Sequence 41, Appl
30	11	55.0	25	1	US-07-786-902-11	Sequence 11, Appl
31	11	55.0	27	2	US-08-758-306-190	Sequence 190, App
32	11	55.0	32	1	US-08-330-535A-23	Sequence 23, Appl
33	11	55.0	32	2	US-08-688-145-20	Sequence 20, Appl
34	11	55.0	32	3	US-08-838-844-23	Sequence 23, Appl
35	11	55.0	34	2	US-08-170-290A-27	Sequence 27, Appl
36	11	55.0	35	1	US-08-125-618-8	Sequence 8, Appl
37	11	55.0	35	4	US-08-397-504-8	Sequence 8, Appl
c 38	11	55.0	39	3	US-08-797-689-29	Sequence 29, Appl
c 39	11	55.0	46	1	US-07-958-140-5	Sequence 5, Appl
c 40	11	55.0	46	6	PCT-US93-09166-5	Sequence 5, Appl
41	11	55.0	53	3	US-08-842-842-3	Sequence 3, Appl
42	11	55.0	79	2	US-08-447-169A-72	Sequence 72, Appl
43	11	55.0	79	3	US-08-233-012C-72	Sequence 72, Appl
44	10.8	54.0	80	4	US-08-477-527A-103	Sequence 103, App
45	10.8	54.0	80	5	US-08-481-710-103	Sequence 103, App

#### ALIGNMENTS

RESULT 1  
US-08-850-049-18  
; Sequence 18, Application US/08850049  
; Patent No. 5965726  
; GENERAL INFORMATION:  
; APPLICANT:  
; APPLICANT:  
; TITLE OF INVENTION: METHOD OF ELIMINATING  
; TITLE OF INVENTION: INHIBITORY/INSTABILITY REGIONS OF MRNA  
; NUMBER OF SEQUENCES: 130  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/850,049  
; FILING DATE: 02-MAY-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/050,478  
; FILING DATE: 26-OCT-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/02908  
; FILING DATE: 29-MAR-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/858,747  
; FILING DATE: 27-MAR-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MORRIS, MARY J.  
; REGISTRATION NUMBER: 34,398  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)758-4800  
; TELEFAX: (212)751-6849  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 50 BASE PAIRS  
; TYPE: NUCLEIC ACID





Query Match	59.0%;	Score 11.8;	DB 5;	Length 68;
Best Local Similarity	66.7%;	Pred. NO. 4e+02;		
Matches 10; Conservative	3;	Mismatches 2;	Indels 0;	Gaps 0;

```

Qy      6 ygcncarcntaygt 2
          :||| ||:|| ||:||
Db      20 TGGCGAGCCCTACGT 6

```

RESULT	5
PCT-US95-01185-99	

```
; Sequence 99, Application PC/TUS9501185
; GENERAL INFORMATION:
```

Query Match	59.0%;	Score 11.8;	DB 6;	length 68;
Best Local Similarity	66.7%;	Pred. No. 4e+02;		
Matches 10;	Conservative 3;	Mismatches 2;	Indels 0;	Gaps 0;

QY	6	ygncarcntaygt	20
		:   :   :	
Db	53	TGGCGAGCCCTACGT	67

RESULT 6  
PCT-US95-01185-100/c

```

; Sequence 100, Application PC/TUS95011855
; GENERAL INFORMATION:

```

```

:      TYPE: nucleic acid
:      STRANDEDNESS: single
:      TOPOLOGY: linear
:
:      MOLECULE TYPE: other nucleic acid
:      DESCRIPTION: /desc = "synthetic DNA"
PCT-US95-01185-100

```

Query Match	59.0%;	Score 11.8;	DB 6;	Length 68;
Best Local Similarity	66.7%;	Pred. No. 4e+02;		
Matches 10; Conservative	3;	Mismatches 2;	Indels 0;	Gaps 0;

```
QY      6 ygcncarcntaygt 20
          :|||:|||||
Db      20 TGGCAGCCCTACGT 6
```

RESULT 7  
US-07-734-

; Sequence 37, Application US/07734225A  
; Patent No. 5320840

Query Match 58.0%; Score 11.6; DB 1; Length 30;  
Best Local Similarity 64.7%; Pred. No. 4.8e+02;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcngaygcncarcnta 17  
|||:|:|:|:|:|:|  
DB 12 GCAGACGCGCAGCGCTA 28

## RESULT 8

US-07-692-995B-37  
; Sequence 37, Application US/07692995B  
; Patent No. 5416195  
; GENERAL INFORMATION:  
; APPLICANT: Camble, Roger  
; APPLICANT: Heather, Carr  
; APPLICANT: Timms, David  
; APPLICANT: Wilkinson, Anthony J.  
; TITLE OF INVENTION: POLYPEPTIDES  
; NUMBER OF SEQUENCES: 66  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
; STREET: Eleventh Floor, 1615 L. Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/692,995B  
; FILING DATE: 19910429  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9009623.1  
; FILING DATE: 30-APR-1990  
; APPLICATION DATA:  
; APPLICATION NUMBER: GB 9013773.8  
; FILING DATE: 20-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9016215.7  
; FILING DATE: 24-JUL-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9102799.5  
; FILING DATE: 11-FEB-1991  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 861-3000  
; TELEFAX: (202) 822-0944  
; TELEX: 248453 CUSH  
; INFORMATION FOR SEQ ID NO: 37:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-07-692-995B-37

Query Match 58.0%; Score 11.6; DB 1; Length 30;  
Best Local Similarity 64.7%; Pred. No. 4.8e+02;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcngaygcncarcnta 17  
|||:|:|:|:|:|:|  
DB 12 GCAGACGCGCAGCGCTA 28

## RESULT 9

US-08-488-457-37  
; Sequence 37, Application US/08488457  
; Patent No. 5773581  
; GENERAL INFORMATION:  
; APPLICANT: Camble, Roger  
; APPLICANT: Timms, David  
; APPLICANT: Wilkinson, Anthony J.  
; TITLE OF INVENTION: CONTINUOUS RELEASE PHARMACEUTICAL  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
; STREET: Eleventh Floor, 1615 L. Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036-5601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/488,457  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/155,327  
; FILING DATE: 22-NOV-1993  
; APPLICATION NUMBER: US 07/734,225  
; FILING DATE: 22-JUL-1991  
; APPLICATION NUMBER: GB 9016138.1  
; FILING DATE: 23-JUL-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9018414.4  
; FILING DATE: 23-AUG-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9018415.1  
; FILING DATE: 23-AUG-1990  
; APPLICATION DATA:  
; APPLICATION NUMBER: GB 9018416.9  
; FILING DATE: 23-AUG-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9018417.1  
; FILING DATE: 23-AUG-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9018418.5  
; FILING DATE: 23-AUG-1990  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)861-3000  
; TELEFAX: (202)822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 37:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-488-457-37

Query Match 58.0%; Score 11.6; DB 2; Length 30;  
Best Local Similarity 64.7%; Pred. No. 4.8e+02;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcngaygcncarcnta 17  
|||:|:|:|:|:|:|  
DB 12 GCAGACGCGCAGCGCTA 28

## RESULT 10

US-08-366-953A-18  
; Sequence 18, Application US/08366953A



NAME: Barry J. Swanson  
REGISTRATION NUMBER: 33,215  
REFERENCE/DOCKET NUMBER: NEX40-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 793-3333  
TELEFAX: (303) 793-3433  
INFORMATION FOR SEQ ID NO: 89:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 71 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: RNA  
FEATURE:  
OTHER INFORMATION: All C's are 2'-NH2 cytosine  
FEATURE:  
OTHER INFORMATION: All U's are 2'-NH2 uracil  
US-08-479-724A-89

Query Match 58.0%; Score 11.6; DB 2; Length 71;  
Best Local Similarity 57.9%; Pred. No. 5.2e+02;  
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 cngaygcncarcntaygt 20  
| ||||| : | ||||  
DB 36 CAGACGCGCTACTTTATGT 18

RESULT 13  
US-08-472-256B-89/C  
Sequence 89, Application US/08472256B  
Patent No. 6001988  
GENERAL INFORMATION:  
APPLICANT: PARMA, DAVID  
TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS  
TITLE OF INVENTION: TO LECTINS  
NUMBER OF SEQUENCES: 177  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Swanson & Bratschun, L.L.C.  
STREET: 8400 E. Prentice Avenue, Suite 200  
CITY: Englewood  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MB  
COMPUTER: IBM pc compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,256B  
FILING DATE: 07-JUNE-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/714,131  
FILING DATE: 10-JUNE-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/536,428  
FILING DATE: 11-JUNE-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/964,624  
FILING DATE: 21-OCTOBER-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Barry J. Swanson  
REGISTRATION NUMBER: 33,215  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 793-3333  
TELEFAX: (303) 793-3433  
INFORMATION FOR SEQ ID NO: 89:  
SEQUENCE CHARACTERISTICS:

LENGTH: 71 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: RNA  
FEATURE:  
OTHER INFORMATION: All C's are 2'-NH2 cytosine  
FEATURE:  
OTHER INFORMATION: All U's are 2'-NH2 uracil  
US-08-472-256B-89

Query Match 58.0%; Score 11.6; DB 5; Length 71;  
Best Local Similarity 57.9%; Pred. No. 5.2e+02;  
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 cngaygcncarcntaygt 20  
| ||||| : | ||||  
DB 36 CAGACGCGCTACTTTATGT 18

RESULT 14  
PCT-US96-09455A-89/C  
Sequence 89, Application PC/TUS9609455A  
GENERAL INFORMATION:  
APPLICANT: PARMA, et al.  
TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID  
TITLE OF INVENTION: LIGANDS TO LECTINS  
NUMBER OF SEQUENCES: 390  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Swanson & Bratschun, L.L.C.  
STREET: 8400 E. Prentice Avenue, Suite 200  
CITY: Englewood  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MB  
COMPUTER: IBM pc compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/09455A  
FILING DATE: 05 JUNE 1996  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/479,724  
FILING DATE: 07-JUNE-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/472,256  
FILING DATE: 07-JUNE-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/472,255  
FILING DATE: 07-JUNE-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/477,829  
FILING DATE: 07-JUNE-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Barry J. Swanson  
REGISTRATION NUMBER: 33,215  
REFERENCE/DOCKET NUMBER: NEX40C/PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 793-3333  
TELEFAX: (303) 793-3433  
INFORMATION FOR SEQ ID NO: 89:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 71 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: RNA  
FEATURE:  
OTHER INFORMATION: All C's are 2'-NH2 cytosine

FEATURE:  
OTHER INFORMATION: All U's are 2'-NH2 uracil  
PCT-US96-09455A-89

## Query Match

58.0%; Score 11.6; DB 6; Length 71;

Best Local Similarity 57.9%; Pred. No. 5.2e+02;

Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 cngaygcncarcntaygt 20

DB 36 CAGACGCGCTACTTATGT 18

## RESULT 15

US-08-852-807-28/c

Sequence 28, Application US/08852807

Patent No. 5861298

GENERAL INFORMATION:

APPLICANT: Debouck, Christine

APPLICANT: Drake, Fred

APPLICANT: Gowen, Maxine

APPLICANT: Rood, Julie

APPLICANT: Hastings, Gregg

APPLICANT: Adams, Mark

APPLICANT: Fraser, Claire

APPLICANT: Lee, No. 5861298man

APPLICANT: Kirkness, Ewen

APPLICANT: Fitzgerald, Lisa

TITLE OF INVENTION: CATHEPSIN K GENE

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: Smithkline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406-2799

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/852,807

FILING DATE: 07-MAY-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/019,942

FILING DATE: 14-JUNE-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/020,273

FILING DATE: 17-JUNE-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/026,273

FILING DATE: 26-AUG-1996

ATTORNEY/AGENT INFORMATION:

NAME: Han, William T

REGISTRATION NUMBER: 34,344

REFERENCE/DOCKET NUMBER: ATG50006-2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-5219

TELEFAX: 610-270-5090

TELEX:

INFORMATION FOR SEQ. ID NO: 28:

SEQUENCE CHARACTERISTICS:

LENGTH: 21 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

HYPOTHETICAL: NO

ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
US-08-852-807-28

## Query Match

57.0%; Score 11.4; DB 3; Length 21;

Best Local Similarity 60.0%; Pred. No. 6e+02;

Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 gcnaygcncarcntaygt 20

DB 21 GAAGATGCGCTACCATATGT 2

Search completed: June 4, 2000, 16:09:39  
Job time: 28058 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:05:28 ; Search time 1236.38 Seconds

(Without alignments)  
-15.736 Million cell updates/sec

Title: US-09-164-714-5

Perfect score: 20  
Sequence: 1 atnccrtaengcngtrtlytt 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 segs, -486395729 residues

Total number of hits satisfying chosen parameters: 356616

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pt1:\*  
10: gb\_pt2:\*  
11: gb\_pt3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vl:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_or:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl1:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vl:\*  
32: gb\_hg1:\*  
33: gb\_hg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pt4:\*  
41: gb\_hg3:\*  
42: gb\_hg4:\*  
43: gb\_hg5:\*  
44: gb\_hg6:\*

45: gb\_hg7:\*  
46: em\_hg1:\*  
47: em\_hg2:\*  
48: em\_hg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pt5:\*  
52: gb\_hg8:\*  
53: gb\_hg9:\*  
54: gb\_hg10:\*  
55: gb\_hg11:\*  
56: gb\_hg12:\*  
57: gb\_hg13:\*  
58: gb\_hg14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query length	DB ID	Description
1	12.6	63.0	91	3 BTU19470	U19470 Bos taurus
2	12.6	63.0	91	3 BTU19472	U19472 Bos taurus
3	12.6	63.0	91	3 CDU19479	U19479 Dama dama 1
4	12.6	63.0	91	3 CHU19477	U19477 Capra hircu
5	12.6	63.0	91	3 OAU19474	U19474 Ovis aries
6	12.6	63.0	91	3 OAU19475	U19475 Ovis aries
7	12.6	63.0	97	5 AR030732	AR030732 Sequence
8	12.2	61.0	48	5 113041	113041 Sequence 18
9	12.2	61.0	77	5 140687	140687 Sequence 18
10	12	60.0	49	5 AR061917	AR061917 Sequence
11	12	60.0	49	5 E06680	E06680 DNA probe t
12	12	60.0	65	5 A87799	A87799 Sequence 18
13	12	60.0	65	5 A87800	A87800 Sequence 19
14	11.8	59.0	61	5 E17241	E17241 Primer. 7/1
15	11.8	59.0	87	5 AR003822	AR003822 Sequence
16	11.8	59.0	87	5 AR010158	AR010158 Sequence
17	11.8	59.0	87	5 AR034690	AR034690 Sequence
18	11.8	59.0	87	5 AR048571	AR048571 Sequence
19	11.8	59.0	87	5 AR055364	AR055364 Sequence
20	11.8	59.0	87	5 140575	140575 Sequence 13
21	11.6	58.0	44	5 113034	113034 Sequence 11
22	11.6	58.0	48	5 113042	113042 Sequence 19
23	11.6	58.0	55	7 XSAMTEO3SB	D16219 Candida alb
24	11.6	58.0	91	3 BTU19471	U19471 Bos taurus
25	11.4	57.0	29	5 144871	144871 Sequence 8
26	11.4	57.0	40	5 AR029226	AR029226 Sequence
27	11.4	57.0	80	9 HSU14601	U14601 Homo sapien
28	11.4	57.0	80	9 HUMSINE	L16226 Human SINE
29	11.2	56.0	39	5 AR031240	AR031240 Sequence
30	11.2	56.0	48	5 AR031254	AR031254 Sequence
31	11.2	56.0	81	16 AF040873	AF040873 Hepatitis
32	11.2	56.0	91	12 ABO10315	ABO10315 Mus muscu
33	11.2	55.0	27	5 140766	140766 Sequence 97
34	11	55.0	29	5 AR024157	AR024157 Sequence
35	11	55.0	33	5 A69052	A69052 Sequence 10
36	11	55.0	60	8 CNS01A9T	A112665 Botrytis
37	11	55.0	60	12 S63846S1	S63846 dopamine D3
38	11	55.0	75	10 HSNCDX10D	Z27430 H.sapiens N
39	11	55.0	78	10 HSA243199	AJ243199 Homo sapi
40	11	55.0	83	4 XU72860	U72860 Xenopus lae
41	11	55.0	91	3 CDU19480	U19480 Dama dama 1
42	11	55.0	91	3 CHU19476	U19476 Capra hircu
43	11	55.0	91	3 CHU19478	U19478 Capra hircu
44	11	55.0	91	3 OAU19473	U19473 Ovis aries
45	10.8	54.0	20	5 E08796	E08796 PCR primer

#### ALIGNMENTS

RESULT 1  
BTU19470/c 91 bp DNA MAM 24-JAN-1996  
LOCUS Bos taurus lysozyme gene, isolate K1, partial exon 2 and partial  
DEFINITION cds.  
ACCESSION U19470  
VERSION U19470.1 GI:841216  
KEYWORDS  
SOURCE  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
Bovinae; Bos.  
REFERENCE 1 (bases 1 to 91)  
AUTHORS Irwin,D.M.  
TITLE Evolution of the bovine lysozyme gene family: changes in gene  
expression and reversal of function  
JOURNAL J. Mol. Evol. 41 (3), 299-312 (1995)  
MEDLINE 96054033  
BASE COUNT 2 (bases 1 to 91)  
ORIGIN Irwin,D.M.

FEATURES  
source Location/Qualifiers  
1..91  
/organism="Bos taurus"  
/isolate="K1"  
/db\_xref="taxon:9913"  
<1..>91  
/number=2  
<1..>91  
/codon\_start=2  
/product="lysozyme"  
/protein\_id="AA05545.1"  
/db\_xref="GI:841217"  
/translation="WESNYNTRTNYNNGKSDYGIPOINSRW"  
BASE COUNT 36 a 18 c 18 g 19 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 3; Length 91;  
Best Local Similarity 60.0%; Pred. No. 1.2e+04;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 atnccrtangcngtrtlytt 20  
||||:| | | | |  
Db 69 ATCCCATATCATGCTTTT 50

RESULT 2  
BTU19472/c 91 bp DNA MAM 24-JAN-1996  
LOCUS Bos taurus lysozyme gene, isolate K4, partial exon 2 and partial  
DEFINITION cds.  
ACCESSION U19472  
VERSION U19472.1 GI:841219  
KEYWORDS  
SOURCE  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
Bovinae; Bos.  
REFERENCE 1 (bases 1 to 91)  
AUTHORS Irwin,D.M.  
TITLE Evolution of the bovine lysozyme gene family: changes in gene  
expression and reversal of function  
JOURNAL J. Mol. Evol. 41 (3), 299-312 (1995)  
MEDLINE 96054033  
BASE COUNT 2 (bases 1 to 91)  
ORIGIN Irwin,D.M.

FEATURES  
source Location/Qualifiers  
1..91  
/organism="Bos taurus"  
/isolate="K4"  
/db\_xref="taxon:9913"  
<1..>91  
/number=2  
<1..>91  
/codon\_start=2  
/product="lysozyme"  
/protein\_id="AA05545.1"  
/db\_xref="GI:841234"  
/translation="WESNYNTRTNYNNGKSDYGIPOINSRW"  
BASE COUNT 37 a 19 c 18 g 17 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 3; Length 91;

TITLE Direct Submission  
JOURNAL Submitted (04-JAN-1995) David M. Irwin, Clinical Biochemistry,  
University of Toronto, 100 College St., Toronto, Ont M5G 1L5,  
Canada  
FEATURES  
source Location/Qualifiers  
1..91  
/organism="Bos taurus"  
/isolate="K4"  
/db\_xref="taxon:9913"  
<1..>91  
/number=2  
<1..>91  
/codon\_start=2  
/product="lysozyme"  
/protein\_id="AA05546.1"  
/db\_xref="GI:841220"  
/translation="WESNYNTRATNYNNGKSDYGIPOINSHW"  
BASE COUNT 36 a 17 c 18 g 20 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 3; Length 91;  
Best Local Similarity 60.0%; Pred. No. 1.2e+04;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 atnccrtangcngtrtlytt 20  
||||:| | | | |  
Db 69 ATCCCATATCATGCTTTT 50

RESULT 3  
CDU19479 91 bp DNA MAM 24-JAN-1996  
LOCUS Dama dama lysozyme gene, isolate K2, partial exon 2 and partial  
DEFINITION cds.  
ACCESSION U19479  
VERSION U19479.1 GI:841233  
KEYWORDS  
SOURCE  
ORGANISM fallow deer.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervidae; Cervidae;  
Cervinae; Cervus.  
REFERENCE 1 (bases 1 to 91)  
AUTHORS Irwin,D.M.  
TITLE Evolution of the bovine lysozyme gene family: changes in gene  
expression and reversal of function  
JOURNAL J. Mol. Evol. 41 (3), 299-312 (1995)  
MEDLINE 96054033  
BASE COUNT 2 (bases 1 to 91)  
AUTHORS Irwin,D.M.  
TITLE Direct Submission  
JOURNAL Submitted (04-JAN-1995) David M. Irwin, Clinical Biochemistry,  
University of Toronto, 100 College St., Toronto, Ont M5G 1L5,  
Canada  
FEATURES  
source Location/Qualifiers  
1..91  
/organism="Cervus dama"  
/isolate="K2"  
/db\_xref="taxon:30532"  
<1..>91  
/number=2  
<1..>91  
/codon\_start=2  
/product="lysozyme"  
/protein\_id="AA05553.1"  
/db\_xref="GI:841234"  
/translation="WESNYNTRQATNYNNGKSDYGIPOINSRW"  
BASE COUNT 37 a 19 c 18 g 17 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 3; Length 91;



```
CDS
<1..>91
/codon_start=2
/product="lysozyme"
/protein_id="AA085549.1"
```

BASE COUNT 35 a 17 c 20 g 19 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 3; Length 91;  
Best Local Similarity 60.0%; Pred. No. 1.2e+04;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 atnccrtangcngtrtlyt 20  
Db 69 ATCCCATATCATGCTGCTTT 50

RESULT 7  
LOCUS AR030732 97 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 17 from patent US 5861298.  
ACCESSION AR030732  
VERSION AR030732.1 GI:5943946  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE  
AUTHORS 1 (bases 1 to 97)  
Adams,M.D., Blake,J.A., Debouck,C.M., Drake,F.H., Fitzgerald,L.M.,  
Fraser,C.M., Gowen,M., Hastings,G.A., Kirkness,E.F., Lee,N.H. and  
Rood,J.  
TITLE Cathepsin K gene  
JOURNAL Patent: US 5861298-A 17 19-JAN-1999;  
FEATURES  
source Location/Qualifiers  
1..97  
/organism="unknown"  
BASE COUNT 31 a 24 c 25 g 17 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 5; Length 97;  
Best Local Similarity 60.0%; Pred. No. 1.2e+04;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 atnccrtangcngtrtlyt 20  
Db 69 ATGCCACAGCGCTGTCTT 50

RESULT 8  
LOCUS I13041 48 bp DNA PAT 26-JUL-1995  
DEFINITION Sequence 18 from patent US 5432270.  
ACCESSION I13041  
VERSION I13041.1 GI:910424  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE  
AUTHORS 1 (bases 1 to 48)  
Zaslloff,M.A., Beylins,C.L. and Diamond,G.  
TITLE DNA encoding tracheal antimicrobial peptides  
JOURNAL Patent: US 5432270-A 18 11-JUL-1995;  
FEATURES  
source Location/Qualifiers  
1..48  
/organism="unknown"  
BASE COUNT 14 a 14 c 8 g 12 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 5; Length 48;  
Best Local Similarity 61.1%; Pred. No. 2e+04;  
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 atnccrtangcngtrtlyt 18

Db 47 ATCCGTATGAGCTGTGTT 30

RESULT 9  
LOCUS I40687 77 bp DNA PAT 13-MAY-1997  
DEFINITION Sequence 18 from patent US 5622828.  
ACCESSION I40687  
VERSION I40687.1 GI:2082167  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE  
AUTHORS 1 (bases 1 to 77)  
Parma,D.H. and Gold,L.  
TITLE High-affinity oligonucleotide ligands to secretory phospholipase A2  
(sPLA<sub>2</sub> sub 2)  
JOURNAL Patent: US 5622828-A 18 22-APR-1997;  
FEATURES  
source Location/Qualifiers  
1..77  
/organism="unknown"  
BASE COUNT 22 a 19 c 19 g 17 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 5; Length 77;  
Best Local Similarity 61.1%; Pred. No. 2e+04;  
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 2 tncctangcngtrtlyt 19  
Db 35 TACCGTAAGACGTATTC 52

RESULT 10  
LOCUS AR061917 49 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 2 from patent US 5843667.  
ACCESSION AR061917  
VERSION AR061917.1 GI:5989608  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE  
AUTHORS 1 (bases 1 to 49)  
Weisburg,W.G. and Pelletier,D.A.  
TITLE Nucleic acid probes for the detection for genital mycoplasmas  
JOURNAL Patent: US 5843667-A 2 01-DEC-1996;  
FEATURES  
source Location/Qualifiers  
1..49  
/organism="unknown"  
BASE COUNT 11 a 13 c 8 g 17 t  
ORIGIN

Query Match 60.0%; Score 12; DB 5; Length 49;  
Best Local Similarity 70.6%; Pred. No. 2.6e+04;  
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 atnccrtangcngtrtlyt 17  
Db 33 ATGCCCTATGCGGTATT 49

RESULT 11  
LOCUS E06680 49 bp DNA PAT 29-SEP-1997  
DEFINITION DNA probe to detect Mycoplasma hominis, ureaplasma urealyticum and  
Mycoplasma genitalium.  
ACCESSION E06680  
VERSION E06680.1 GI:2174862  
KEYWORDS JP 1994030800-A/2.

ACCESSION	A87800
VERSION	A87800.1
KEYWORDS	GI:67369391
SOURCE	.
ORGANISM	human.
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Euarchonta; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 65)
AUTHORS	Karayam,I., and Hong,S.-S.
TITLE	USE OF A POLYPEPTIDE AS CELL RECEPTOR FOR ADENOVIRUSES
JOURNAL	Patient: WO 9833929-A 06-AUG-1998;
FEATURES	KARAYAM LUCIE (FR); HONG SAW SEE (FR) Location/Qualifiers
source	. .65 /organism="Homo sapiens" /isolate="OLIGONUCLEOTIDE DE SYNTHÈSE CODANT POUR AA 157-176 DU MHC-I"
BASE COUNT	/db_xref="taxon:9606" 20 a 27 c 6 g 12 t
ORIGIN	
Query Match	Score 60.0%; DB 5; Length 65; Best Local Similarity 63.2%; Pred. No. 2,6e+04; Matches 12; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY	1 atnccrtangcngtfttyt 19            :   :
Dd	38 ATGGCTTAGCGGATATT 20
RESULT 14	
EI7241/C	
LOCUS	EI7241 61 bp DNA PAT 28-JUL-1999
DEFINITION	Primer.
ACCESSION	EI7241
VERSION	EI7241.1 GI:5711924
KEYWORDS	JP 1998257893-A/18.
SOURCE	unidentified.
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 61)
AUTHORS	Nakamura,K. and Hanai,N..
TITLE	HOMAN COMPLEMENTARTY DETERMINING REGION(CDR) TRANSPLANTED ANTIBODY
JOURNAL	AGAINST GANGLIOSIDE GM2 Patent: JP 1998257893-A 29-SEP-1998;
COMMENT	KYOMA HAKKO KOCYO CO LTD OS None OC Artificial sequences. PN JP 1998257893-A/18 PD 29-SEP-1998 PF 19-MAR-1997 JP 1997066981 PI NAKAMURA KAZUYASU, HANAII NOBUO PC C12N15/09,A6IK39/395,A6IK39/395,C07K16/10,C12N5/10,C12P21/08, PC G01M33/531,G01N33/574,G01N33/577,(C12N5/10,C12R1:91), PC (C12P21/08,C12R1:91); CC strandedness: Single; CC topology: Linear; CC hypothetical: No; FH key Location/Qualifiers FT source . .61 location/artificial sequences'. . .61 location/qualifiers /organism='unidentified' /db_xref='taxon:32644'
BASE COUNT	18 a 24 c 14 g 5 t
ORIGIN	
Query Match	Score 59.0%; DB 5; Length 61;



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:24:01 ; Search time 322.35 Seconds  
(without alignments)  
15.523 Million cell updates/sec

Title: US-09-164-714-5

Perfect score: 20  
Sequence: 1 atnccrtaangcngtrtytt 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 431286

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12.6	63.0	97	1	V09676 Human cathepsin K
2	12.2	61.0	48	1	Oligo probe TAP160
3	12.2	61.0	77	1	Secretory phosphol
4	12.0	60.0	49	1	Nucleic acid probe
5	12.0	60.0	49	1	U. urealyticum 16S
6	12.0	60.0	59	1	Synthetic Fns gene
7	11.8	59.0	61	1	Primer for coding
8	11.8	59.0	80	1	De-immunised 708 V
9	11.8	59.0	87	1	HUN-G15 Antibodie
10	11.8	59.0	87	1	Humanised H65 heav
11	11.6	58.0	44	1	TAP probe (1). New
12	11.6	58.0	44	1	Probe for bovine t
13	11.6	58.0	48	1	Oligo probe TAP113
14	11.4	57.0	29	1	Salmonella enterit
15	11.4	57.0	34	1	BH3 interacting do
16	11.4	57.0	40	1	RF-X polynucleotid
17	11.4	57.0	58	1	Human gene signatu
18	11.2	56.0	21	1	Primer R1 for H.py
19	11.2	56.0	62	1	L-selectin orphan
20	11.2	56.0	67	1	Staphylococcus aur
21	11.2	55.0	23	1	nachr beta2 subuni
22	11.1	55.0	27	1	Secretory phosphol
23	11.1	55.0	29	1	Herpes simplex vir
24	11.1	55.0	30	1	Oligonucleotide ZC
25	11.1	55.0	31	1	Nucleotide fragmen
26	11.1	55.0	33	1	Primer AB050 for C
27	11.1	55.0	51	1	Murine osteoproteg
28	11.1	55.0	53	1	Staphylococcus aur
29	11.1	55.0	62	1	Downstream sequenc
30	11.1	55.0	88	1	Human gene signatu
31	10.8	54.0	100	1	Downstream sequenc
32	10.8	54.0	20	1	Lactobacillus sp.
33	10.8	54.0	31	1	Human biallelic po
34	10.8	54.0	68	1	Embryonic DNA sequ

35	10.8	54.0	79	1	003346	Sequence from chro
36	10.6	53.0	19	1	X07468	Human B5124 specif
37	10.6	53.0	22	1	082111	Chromosome 11 (loc
38	10.6	53.0	24	1	T10614	LN clone E2 Vbeta8
39	10.6	53.0	26	1	V41471	Nucleotide sequenc
40	10.6	53.0	29	1	Q76064	hPV Primer set 13,
41	10.6	53.0	31	1	084448	Mycobacterium marl
42	10.6	53.0	31	1	V15800	Primer for wild ty
43	10.6	53.0	31	1	X06424	Human biallelic po
44	10.6	53.0	32	1	T30768	Yeast his3 gene PC
45	10.6	53.0	40	1	V08300	Primer for HIV-1 p

## ALIGNMENTS

```
RESULT 1
V09676/c
ID V09676 standard: DNA; 97 BP.
AC V09676;
DT 20-JUL-1998 (first entry) 8.
DE Human cathepsin K gene exon 8.
KW Cathepsin K; human; osteoporosis; periodontal disease;
KW Paget's disease; Gaucher's disease; Alzheimer's disease;
KW central nervous system inflammation; hyperparathyroidism;
KW bone degradation; dental implant degradation; metastasis; tumour;
KW diagnosis; therapy; ss.
OS Homo sapiens.
PN EP-812916-A2.
PD 17-DEC-1997.
PF 19-MAY-1997; 303395.
PR 26-AUG-1996; US-026083.
PR 14-JUN-1996; US-019942.
PR 17-JUN-1996; US-020273.
PR 26-AUG-1996; WO-014026.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (GENO-) INST GENOMIC RES.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Adams MD, Blake JA, Dabouck CM, Drake FH, Fitzgerald LM,
PI Fraser CM, Gowen M, Hastings GA, Kirkness EF, Lee NH,
PI Rioud J;
PI WPI: 98-034977/04.
DR P-PSDB: W39270.
PT DNA encoding human cathepsin K - useful for diagnosing and treating
PT diseases associated with cathepsin K e.g. osteoporosis, bone
PT degradation, metastatic tumours, etc
PS Claim 13: Page 45: 84pp: English.
CC This genomic DNA sequence comprises the exon 8 of the human
CC cathepsin K gene (see V09660). It encodes a 32-amino acid
CC C-terminal peptide (see W39270) of cathepsin K (see W39216).
CC Polymorphisms in the 8 exons of the cathepsin K gene serve as
CC markers for osteoporosis, periodontal disease, Paget's disease,
CC Gaucher's disease, central nervous system inflammation, Alzheimer's
CC disease, hyperparathyroidism, bone degradation, metastatic tumours,
CC and degradation of bone implants and prostheses, especially dental
CC implants. Ribozymes based e.g. on splice junctions of cathepsin K
CC gene polynucleotides can be used to inhibit cathepsin K gene
CC expression. Isolated polypeptides encoded by exons 1, 2, 3, 4, 5,
CC 6, 7 and 8 are claimed (see W39264-70), as well as those encoded by
CC exon pairs 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 2-4, 2-5, 2-6, 2-7, 2-8,
CC 3-4, 3-5, 3-6, 3-7, 3-8, 4-5, 4-6, 4-7, 4-8, 5-7, 5-8 and 6-8.
CC They can be used to treat a patient in need of cathepsin K or to
CC identify compounds which bind to and inhibit activation of
CC cathepsin K.
SQ Sequence 97 BP; 31 A; 24 C; 25 G; 17 T;
```

Query Match 63.0%; Score 12.6; DB 1; Length 97;  
Best Local Similarity 60.0%; Pred. No. 2.2e+02;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

1 atnccrtaangcngtrtytt 20  
|||:||||:|

DB 69 ATGCCACAGCGCTGTCTT 50

RESULT 2

079194/c

ID 079194 standard; DNA; 48 BP.

AC 079194;

DT 04-AUG-1995 (first entry)

DE Oligo probe TAP160a for bovine tracheal antimicrobial peptide gene.

KW Antimicrobial peptide; TAP; probe; ss.

OS Synthetic.

PN WO9426106-A.

PD 24-NOV-1994.

PF 11-MAY-1994; 005257.

PR 11-MAY-1993; US-060822.

PA (CHIL-) CHILDRENS HOSPITAL PHILADELPHIA.

PI Bevins CL, Diamond G, Zasloff MA;

DR WPI. 95-006233/01.

PT New antimicrobial peptide and precursor from mammalian trachea -

PR Pharmaceuticals and disinfectants

PS Example; page 27; 69 pp; English.

CC A mammalian TAP (R66203) was isolated by a combination of

CC size-exclusion, ion-exchange, and reverse-phase chromatographic

CC fractionations using antimicrobial activity against a strain of

CC E.coli as a functional assay. The mol. wt. of TAP is 4085 Da. The

CC predicted PI is 13.0 and there are no aromatic residues. A cDNA

CC sequence corresp. to the precursor peptide was cloned (see 079533/

CC R66204), and the genomic TAP gene (079534/R66205) was isolated.

CC PCR primers for amplifying the TAP gene from bovine genomic DNA

CC were TAPG3 and TAPG10a. Oligo probes for hybridisation were

CC TAP48a, TAP54s, TAP160a, TAP13S, TAP18S.

SQ Sequence 48 BP; 14 A; 14 C; 8 G; 12 T;

Query Match 61.0%; Score 12.2; DB 1; Length 48;

Best Local Similarity 61.1%; Pred. No. 3.3e+02;

Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 atnccrtangcngtrttt 18

DB 47 ATCCCTAGAGCTGTCTT 30

RESULT 3

T88489

ID T88489 standard; RNA; 77 BP.

AC T88489;

DT 17-DEC-1997 (first entry)

DE Secretory phospholipase A2 nucleic acid ligand used in diagnosis.

KW Nucleic acid ligand; secretory phospholipase A2; inflammation;

KW anti-inflammatory; rheumatoid arthritis; septic shock; VEGF;

KW vascular endothelial growth factor; angiogenesis; psoriasis;

KW metastasis; neoplasia; haemangioma; human immunodeficiency virus;

KW HIV; diagnosis; treatment; specific; high affinity; ss.

OS Synthetic.

PN WO9627604-A1.

PD 12-SEP-1996.

PF 04-MAR-1996; 002942.

PR 19-MAY-1995; US-447172.

PR 06-MAR-1995; US-399412.

PR 19-MAY-1995; US-447169.

PA (NEXS-) NEXSTAR PHARM INC.

PI Gold L, Janjic N, Lochrie M, Parma D;

DR WPI. 96-425373/42.

PT New non-naturally occurring nucleic acid ligands to secretory

PT phospholipase A2, VEGF and HIV-1 gag - useful, e.g., for diagnosis

PT and treatment of inflammation, angiogenesis and HIV infection.

PS Claim 9; Page 91; 169pp; English.

CC T88481-T88593 are nucleic acid ligands (NAL) for secretory phospholipase

CC A2 (sPLA2). Other non-naturally occurring ligands for vascular

CC endothelial growth factor (VEGF) and HIV-1 gag were also made. The NAL

CC were identified using the SELEX procedure. They are used as diagnostic

CC or treatment agents. For example, those for sPLA2 are used to study/  
CC diagnose inflammation and as anti-inflammatory agents e.g. for the  
CC treatment of rheumatoid arthritis and septic shock; those for VEGF are  
CC used to treat angiogenesis-mediated diseases e.g. psoriasis, haemangioma,  
CC metastasis; and those for gag are used to inhibit HIV-1 replication.  
CC NAL can also generally be used to identify the corresponding target.  
SQ Sequence 77 BP; 22 A; 19 C; 19 G; 17 U;

Query Match 61.0%; Score 12.2; DB 1; Length 77;

Best Local Similarity 27.8%; Pred. No. 3.5e+02;

Matches 5; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

QY 2 tncctangcngtrttt 19

DB 35 UNCCGAGAGCGAUCUCU 52

RESULT 4

O54447

ID O54447 standard; DNA; 49 BP.

AC O54447;

DT 04-JUL-1994 (first entry)

DE Nucleic acid probe for Ureaplasma urealyticum.

KW Probe; screening; detection; identification; septic abortion;

KW nongonococcal urethritis; pelvic inflammatory disease;

KW postpartum fever; clinical sample; swab; lavage;

KW Mycoplasma genitalium; Mycoplasma hominis; Ureaplasma urealyticum;

OS Synthetic.

PN EP-576742-A.

PD 05-JAN-1994.

PF 04-JUN-1992; 305125.

PR 04-JUN-1992; BP-305125.

PA (STAD ) AMOCO CORP.

PI Pelletier DA, Weisburg WG;

DR WPI. 94-009118/02.

PT Nucleic acid probes for the detection of genital mycoplasma - 1s

PT able to hybridise with RNA or rDNA of pathogenic mycoplasma

PS Claim 9; Page 10; 23pp; English.

CC Probes specific for Ureaplasma urealyticum (O54446-52),

CC Mycoplasma genitalium (O54453-58) and Mycoplasma hominis (O63723-25)

CC can be used for the selection of these three presumptive

CC etiological agents of nongonococcal urethritis, septic abortion,

CC inflammatory disease and postpartum fever from clinical samples such

CC as genital swabs, genital lavage, sputum, throat swabs, blood, urine,

CC cerebrospinal fluid, skin, biopsy, saliva, synovial fluid, bronchial

CC wash, bronchial lavage or other tissue or fluid samples from humans

CC or animals. The probes can also be used to confirm the presence of

CC the genital mycoplasmas in liquid or semi-solid in vitro culture.

CC This probe is specific for the 16S rRNA of Ureaplasma urealyticum.

SQ Sequence 49 BP; 11 A; 13 C; 8 G; 17 T;

Query Match 60.0%; Score 12; DB 1; Length 49;

Best Local Similarity 70.6%; Pred. No. 4.3e+02;

Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 atnccrtangcngtrttt 17

DB 33 ATGCCCTATGCGGTATT 49

RESULT 5

V64974

ID V64974 standard; DNA; 49 BP.

AC V64974;

DT 05-FEB-1999 (first entry)

DE U. urealyticum 16S ribosomal RNA probe 2256.

KW 16S ribosomal RNA; rRNA; probe; detection; etiological agent; PID;

KW nongonococcal urethritis; pelvic inflammatory disease; salpingitis;

OS Synthetic.



CC one of the T-cell epitopes identified in step (b) thereby to eliminate or  
CC reduce the immunogenicity of the protein when exposed to the immune  
CC system of the given species. A method of analysing a pre-existing protein  
CC to predict the basis for immunogenic responses is also provided. The  
CC methods can be used particularly for reducing the immunogenicity of  
CC immunoglobulins or therapeutic proteins, e.g. Streptokinase (SK). The  
CC products can be used for diagnosis and therapy. Sequences V81047-68  
CC represent oligonucleotides used for the construction of de-immunised 708  
CC Vh and V<sub>k</sub>.  
SQ Sequence 80 BP; 24 A; 19 C; 21 G; 16 T;

Query Match 59.0%; Score 11.8; DB 1; Length 80;  
Best Local Similarity 66.7%; Pred. No. 6e+02;  
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 6 rtangengrttytt 20  
:|||:|||:||||  
Db 55 GTRAGGCTGTCTCTT 41

RESULT 9  
044074/c  
ID 044074 standard; DNA; 87 BP.  
AC 044074;  
DT 28-OCT-1993 (first entry)  
DE HUH-G15.  
KW Antibody; variable domain; light; L; heavy; H; PCR;  
KW polymerase chain reaction; H65; monoclonal antibody; MAb; ss.  
OS Synthetic.  
PN WO9311794-A.  
PD 24-JUN-1993.  
PF 14-DEC-1992; 010906.  
PR 13-DEC-1991; US-808464.  
PA (XOMA) XOMA CORP.  
PI Fishwild DM, Kohn FR, Little RG, Studnicka GM;  
DR WPI: 93-213827/26.  
PT Antibodies prepn. used for treatment of auto-immune diseases - by  
PT replacement of critical residues to reduce immunogenicity but  
PT retain binding affinity, etc.  
PS Example 6: Fig 7B; 160pp: English.  
CC Example 6 describes the construction of the gene encoding the  
CC humanised H65 heavy chain contg. the moderate risk residues.  
CC The humanised heavy chain was assembled from the oligonucleotides  
CC given in 044070-75.  
SQ Sequence 87 BP; 22 A; 22 C; 19 G; 24 T;

Query Match 59.0%; Score 11.8; DB 1; Length 87;  
Best Local Similarity 66.7%; Pred. No. 6e+02;  
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 6 rtangengrttytt 20  
:|||:|||:||||  
Db 44 ATRAGCAGTGTCTT 30

RESULT 10  
V31078/c  
ID V31078 standard; DNA; 87 BP.  
AC V31078;  
DT 18-AUG-1998 (first entry)  
DE Humanised H65 heavy chain variable region construction oligo HUH-G15.  
KW Humanised; human; mouse; CD5; anti-CD5 antibody; immunoglobulin;  
KW depletion; cytotoxic; immunocjugate; fusion protein; psoriasis;  
KW autoimmune disease; rheumatoid arthritis; type I diabetes; ss.  
OS Synthetic.  
PN Mus sp.  
OS Homo sapiens.  
PN US5770196-A.  
PD 23-JUN-1998.  
PF 07-JUN-1995; 472788.  
PR 23-JUN-1995; US-082842.

PR 13-DEC-1991; US-808464.  
PR 14-DEC-1992; WO-010906.  
PR 07-JUN-1995; US-472788.  
PA (XOMA) XOMA CORP.  
PI Studnicka GM;  
DR WPI: 98-376744/32.  
PT Depletion of CD5-positive cells in vivo - using anti-CD5 antibodies  
PT with humanised variable regions  
PS Example 7; Fig 7B; 77pp: English.  
CC A method has been developed of depleting CD5+ cells in an animal. The  
CC method comprises administering a cytotoxic protein containing a modified  
CC immunoglobulin (Ig) variable domain, where the protein is an anti-CD5 Ig  
CC molecule or an immunocjugate or fusion protein containing an anti-CD5  
CC Ig molecule, and where the modified Ig variable domain comprises at  
CC least one of (a) a modified light chain variable region (see W58478 or  
CC W58480), and (b) a modified heavy chain variable region (see W58479 or  
CC W58481), where W58478 and W58479 are humanised forms of the H65 light  
CC and heavy chain variable domains with low risk amino acid substitutions  
CC 11.e. low risk of reducing antigen-binding specificity.] and W58480 and  
CC W58481 are humanised forms of the H65 light and heavy chain variable  
CC domains with moderate risk amino acid substitutions and are present in  
CC humanised H65 antibody he3 (ATCC HB 11206). The method is useful for  
CC treating autoimmune diseases, especially systemic lupus erythematosus,  
CC rheumatoid arthritis, psoriasis or type I diabetes. The present sequence  
CC represents an oligonucleotide used in the construction of humanised  
CC anti-CD5 immunoglobulin H65 heavy chain variable region DNA sequence.  
SQ Sequence 87 BP; 22 A; 22 C; 19 G; 24 T;

Query Match 59.0%; Score 11.8; DB 1; Length 87;  
Best Local Similarity 66.7%; Pred. No. 6e+02;  
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 6 rtangengrttytt 20  
:|||:|||:||||  
Db 44 ATRAGCAGTGTCTT 30

RESULT 11  
025193  
ID 025193 standard; DNA; 44 BP.  
AC 025193;  
DT 19-NOV-1992 (first entry)  
DE TAP probe (1).  
KW Tracheal antimicrobial peptide; respiratory mucosa; RFLP; ss.  
OS Synthetic.  
PN WO9207873-A.  
PD 14-MAY-1992.  
PF 24-OCT-1991; 007996.  
PR 25-OCT-1990; US-603451.  
PA (CHIL-) CHILDRENS HOSPITAL PHILADELPHIA.  
PI Bevins CL, Diamond G, Zasloff MA;  
DR WPI: 92-183626/22.  
PT New tracheal antimicrobial peptide and encoding DNA - used for  
PT treating bacterial and fungal infections, used in mouthwashes,  
PT contact disinfectants etc.  
PS Disclosure; Page 22; 42pp: English.  
CC The probes represented in 025193-94 are used in the screening  
CC of bovine tracheal cDNA library.  
SQ Sequence 44 BP; 12 A; 7 C; 12 G; 13 T;

Query Match 58.0%; Score 11.6; DB 1; Length 44;  
Best Local Similarity 64.7%; Pred. No. 7e+02;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 atnrcrtangngtrtt 17  
|||:|||:|||:|  
Db 2 ATCCGTGAAGCTGTGT 18

RESULT 12  
079539



ID Q79539 standard; DNA; 44 BP.  
 AC Q79539;  
 DT 04-AUG-1995 (first entry)  
 DE Probe for bovine tracheal antimicrobial peptide.  
 KW Antimicrobial peptide; TAP; probe; ss.  
 OS Synthetic.  
 PN W09426106-A.  
 PD 24-NOV-1994.  
 PF 11-MAY-1994; U05257.  
 PR 11-MAY-1993; US-060822.  
 PA (CHIL-) CHILDRENS HOSPITAL PHILADELPHIA.  
 PI Bevins CL, Diamond G, Zasloff MA;  
 DR WPI; 95-006233/01.  
 PT New antimicrobial peptide and precursor from mammalian trachea -  
 PT pharmaceuticals and disinfectants  
 PS Example; page 44; 69 pp; English.  
 CC In order to clone TAP cDNA, degenerate oligo primers (Q79535,  
 CC Q79541, Q79536) were designed corresp. to AAs 1-6 (BT-40-1  
 CC and -2) and 21-26 (BT-40-3) respectively. Sense primers 1  
 CC (Q79535) and 2 (Q79541) had the EcoRI recognition site incorporated  
 CC on the 5' end, and antisense primer 3 (Q79536) had the SstI  
 CC recognition site on its 5' end. These primers were used in a PCR  
 CC using bovine tracheal cDNA as template DNA and were expected to  
 CC amplify the nucleotide sequence coding for AAs 1-26. The principle  
 CC DNA product was 90 bp in length, as expected. This indicated the  
 CC peptide was present in the library and yielded a DNA template for  
 CC probe synthesis. The cDNA library from bovine tracheal epithelium  
 CC was screened using three different probes in parallel: the PCR  
 CC product; and two 'best-guess' synthetic oligo probes (Q79538,  
 CC Q79539). The sequence for positive clone pBT40-4.4 is given in  
 CC (see R66204 FT).  
 SQ Sequence 44 BP; 12 A; 7 C; 12 G; 13 T;

Query Match 58.0%; Score 11.6; DB 1; Length 44;  
 Best Local Similarity 64.7%; Pred. No. 7e+02; Indels 0; Gaps 0;

Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 atnccrtangcngtrtlt 17  
 |||:|||||:  
 Db 2 ATCCTGTAGCTGTGT 18

RESULT 13

Q79195

ID Q79195 standard; DNA; 48 BP.  
 AC Q79195;  
 DT 04-AUG-1995 (first entry)  
 DE Oligo probe TAP113s for bovine tracheal antimicrobial peptide gene.  
 KW Antimicrobial peptide; TAP; probe; ss.  
 OS Synthetic.  
 PN W09426106-A.  
 PD 24-NOV-1994.  
 PF 11-MAY-1994; U05257.  
 PR 11-MAY-1993; US-060822.  
 PA (CHIL-) CHILDRENS HOSPITAL PHILADELPHIA.  
 PI Bevins CL, Diamond G, Zasloff MA;  
 DR WPI; 95-006233/01.  
 PT New antimicrobial peptide and precursor from mammalian trachea -  
 PT and related DNA, vector and transformed cells, useful as  
 PT pharmaceuticals and disinfectants  
 PS Example; page 27; 69 pp; English.  
 CC A mammalian TAP (R66203) was isolated by a combination of  
 CC size-exclusion, ion-exchange, and reverse-phase chromatographic  
 CC fractionations using antimicrobial activity against a strain of  
 CC E. coli as a functional assay. The mol. wt. of TAP is 4085 Da. The  
 CC predicted pI is 13.0 and there are no aromatic residues. A cDNA  
 CC sequence corresp. to the precursor peptide was cloned (see Q79533/  
 CC R66204), and the genomic TAP gene (Q79534/R66205) was isolated.  
 CC PCR primers for amplifying the TAP gene from bovine genomic DNA  
 CC were TAP69s and TAP610a. Oligo probes for hybridisation were

CC TAP48a, TAP54s, TAP160a, TAP113s, TAP118s.  
 SQ Sequence 48 BP; 13 A; 8 C; 14 G; 13 T;

Query Match 58.0%; Score 11.6; DB 1; Length 48;  
 Best Local Similarity 64.7%; Pred. No. 7.1e+02;  
 Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 atnccrtangcngtrtlt 17  
 |||:|||||:  
 Db 2 ATCCTGTAGCTGTGT 18

RESULT 14

T74158

ID T74158 standard; DNA; 29 BP.  
 AC T74158;  
 DT 29-SEP-1997 (first entry)  
 DE Salmonella enteritidis tct hybridisation oligonucleotide MKS3.  
 KW Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody;  
 KW ss.  
 OS Synthetic.  
 PN US5635617-A.  
 PD 03-JUN-1997.  
 PF 26-APR-1993; 054452.  
 PR 26-APR-1993; US-054452.  
 PR 26-APR-1994; US-233788.  
 PA (UVVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
 PI Clouthier SC, Collinson SK, Doran JL, Kay MW;  
 DR WPI; 97-309886/28.  
 PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or  
 PT enteropathogenic bacteria of the Enterobacteria family  
 PS Example 6; Column 37; 85pp; English.  
 CC The present sequence represents an oligonucleotide MKS3 that hybridises  
 CC to WK-10, which is used as part of the method in sequencing the tctCBA  
 CC gene cluster of Salmonella enteritidis. The nucleic acids produced can  
 CC be used to provide diagnostic assays for Salmonella and/or  
 CC enteropathogenic bacteria of the family Enterobacteria. They can also  
 CC be used to provide proteins and antibodies which can be used for assays.  
 CC The nucleic acid sequences can be used to provide probes or primers  
 CC which can specifically hybridise to nucleic acid molecules from greater  
 CC than 99% of Salmonella strains that are pathogenic to warm-blooded  
 CC animals relative to nucleic acid molecules from virtually all other  
 CC microbial organisms.  
 SQ Sequence 29 BP; 7 A; 4 C; 2 G; 16 T;

Query Match 57.0%; Score 11.4; DB 1; Length 29;  
 Best Local Similarity 60.0%; Pred. No. 8.6e+02; Indels 0; Gaps 0;

Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 atnccrtangcngtrtlt 20  
 |||:|||||:  
 Db 8 ATTCACACAGCTTTT 27

RESULT 15

V22156/c

ID V22156 standard; cDNA; 34 BP.  
 AC V22156;  
 DT 20-JUL-1998 (first entry)  
 DE BH3 interacting domain death agonist polynucleotide fragment 17.  
 KW Human; BH3 interacting domain death agonist; BID; BCL-2 family;  
 KW apoptosis; regulation; cell death; inflammation; cancer; arthritis;  
 KW autoimmune disease; viral infection; lymphoproliferative; ss.  
 OS Homo sapiens.  
 PN W09809980-A1.  
 PD 12-MAR-1998.  
 PF 09-SEP-1997; U15872.  
 PR 09-SEP-1996; US-706741.  
 PA (UNIM ) UNIV WASHINGTON.  
 PI Korsmeyer SJ;  
 DR WPI; 98-193546/17.

PT BH3 interacting domain death agonist polypeptide - used for treating  
decreased apoptotic conditions resulting from inflammation etc.  
PS Disclosure; Page 23; 118pp; English.  
CC The present sequence represents a BH3 interacting domain death agonist  
(BID) polynucleotide fragment given in the present invention. The  
CC protein, the DNA encoding it or antisense sequences can be used for  
CC preventing or treating a decreased apoptotic state of a cell. The  
CC decreased apoptotic state that is treated results from a disease such as  
CC cancer, viral infections, lymphoproliferative conditions, arthritis,  
CC inflammation and autoimmune diseases. Antibodies against the BID protein  
CC can be used for detecting a BID polypeptide in a cell or population of  
CC cell. The nucleic acid sequence and the BID protein can also be used for  
CC treating immunodeficiency disease (including AIDS), senescence,  
CC neurodegenerative disease, ischemic and reperfusion cell death,  
CC infertility and wound-healing. Primers derived from the nucleic acid  
CC encoding the BID protein can be used for detecting/quantitating the  
CC protein and for detecting alterations in the nucleic acid encoding the  
CC BID protein.  
SQ Sequence 34 BP; 12 A; 9 C; 7 G; 6 T;

## Query Match

57.0%; Score 11.4; DB 1; Length 34;

Best Local Similarity 60.0%; Pred. No. 8.8e+02;

Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 atnccrtangcngtrtttctt 20

DB 22 ATWCCGATGATGTCCTTCTT 3

Search completed: June 4, 2000, 16:24:04  
Job time: 28900 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 13:53:02 ; Search time 4521.53 Seconds  
(without alignments)  
17.929 Million cell updates/sec

Title: US-09-164-714-5

Perfect score: 20

Sequence: 1 atncrtangcgttttctt 20

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 119968

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result

Query

No.	Score	Match	Length	DB	ID	Description
1	12.6	63.0	88	45	AI339161	AI339161 qtl4h01.x
2	12	59.0	72	69	AW116071	AW116071 f106g06.x
3	11.8	59.0	67	23	H41775	H41775 yn95c08.s1
4	11.8	59.0	79	28	AA116248	AA116248 mql1g10.r
5	11.6	58.0	52	51	AT749415	AT749415 at23g11.x
6	11.6	58.0	74	21	T99376	T99376 y64f04.s1
7	11.6	58.0	93	62	AI929034	AI929034 au64c11.y
8	11.6	58.0	94	84	B35774	B35774 HS-1030-BI-
9	11.6	58.0	95	33	AA449805	AA449805 zx09h01.s
10	11.4	57.0	58	20	D25841	D25841 HDMS04213
11	11.4	57.0	80	23	H55643	H55643 CHR220582.C
12	11.4	57.0	84	44	AI318353	AI318353 ta76g04.x
13	11.4	57.0	87	44	AI318677	AI318677 ta76g05.x
14	11.4	57.0	100	40	AA910722	AA910722 ok84e12.s
15	11.4	57.0	100	42	AI109871	AI109871 vg80b10.r
16	11.2	56.0	40	36	AA601989	AA601989 nc08h07.s
17	11.2	56.0	81	79	AW250425	AW250425 2822308.3
18	11.2	56.0	89	45	AI385053	AI385053 fb06e09.x
19	11.2	56.0	91	42	AI130098	AI130098 SMOVL3CAN
20	11.2	56.0	100	37	AA706875	AA706875 z121f09.s
21	11.2	55.0	51	49	AI633437	AI633437 to46h07.x
22	11	55.0	52	49	AT661448	AT661448 va28d11.x
23	11	55.0	64	24	H80115	H80115 yu09g05.s1
24	11	55.0	64	63	AI950653	AI950653 wx51h01.x
25	11	55.0	67	62	AI921937	AI921937 wn86h08.x
26	11	55.0	68	29	AA153886	AA153886 mq50e02.r
27	11	55.0	76	27	AA052289	AA052289 mb42e11.r
28	11	55.0	78	33	AA412487	AA412487 zt99a10.s
29	11	55.0	84	46	AI463806	AI463806 va29d11.y
30	11	55.0	85	28	C21053	C21053 HDMS000255
31	11	55.0	87	50	AI709401	AI709401 as37c10.x
32	11	55.0	88	44	AI318471	AI318471 ta43c04.x
33	11	55.0	88	63	AI956631	AI956631 ul79f04.y
34	11	55.0	89	44	AI318524	AI318524 ta79h04.x
35	11	55.0	90	43	AI204814	AI204814 zf-est134
36	11	55.0	90	45	AI357740	AI357740 qv73c05.x
37	11	55.0	94	44	AI282643	AI282643 qt70a06.x
38	11	55.0	95	29	AA155036	AA155036 ms14e01.r
39	11	55.0	96	46	AI442889	AI442889 sa28c02.x
40	11	55.0	98	37	AA710541	AA710541 vt50h03.r
41	10.8	54.0	43	41	AU006943	AU006943 AU006943
42	10.8	54.0	43	41	AU010101	AU010101 AU010101
43	10.8	54.0	43	41	AU010103	AU010103 AU010103
44	10.8	54.0	58	38	AA767515	AA767515 oa44a03.s
45	10.6	53.0	31	43	AI193015	AI193015 qe69a08.x

## ALIGNMENTS

```
RESULT 1
LOCUS AI339161 88 bp mRNA EST 29-DEC-1998
DEFINITION qtl4h01.x1 NCI-CGAP.GC4 Homo sapiens cDNA clone IMAGE:1947601 3'
SUBMIT 1; mRNA sequence.
ACCESSION AI339161
VERSION AI339161.1 GI:4076088
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 88)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TUMOR National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Tel: (301) 496-1550
Email: Robert.L.Strausberg@nih.gov
```

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/BLNT at:  
www.bio.lnl.gov/dbp/image/image.html

Trace considered overall poor quality  
Seq primer: -40UP from Gibco  
High quality sequence stop: 1.

## FEATURES

source

1..88  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1947601"  
/clone\_1lb="NCI-CGAP.GC4"  
/tissue\_type="pooled germ cell tumors"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker: 1st strand cDNA was prepared from 3 pooled  
germ cell tumors, and was then primed with a Not I -  
oligo(dT) primer. Double-stranded cDNA was ligated to Eco  
RI adaptors (Pharmacia), digested with Not I and cloned  
into the Not I and Eco RI sites of the modified pT7T3  
vector. Library is normalized. Library was constructed by  
Bento Soares and M. Fatima Bonaldo."

## BASE COUNT

19 a 16 c 21 g 32 t

## Query Match

Best local Similarity 63.0%; Score 12.6; DB 45; Length 88;

Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 atnccrtangcgtttt 20

Db 18 atcccgtagcaggttc 37

## RESULT 2

```
LOCUS AW116071 72 bp mRNA EST 20-OCT-1999
DEFINITION f106g06.x1 Sugano Kawakami zebrafish DNA Danio rerio cDNA clone
ACCESSION AW116071
VERSION AW116071.1 GI:6082409
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;  
Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;  
Cyprinidae; Cyprinidae; Rasbora; Danio.

REFERENCE 1 (bases 1 to 72)  
AUTHORS Sugano, S., Kawakami, K., Johnson, S., Li, F., Marra, M., Eddy, S.,  
Hillier, L., Clifton, S., Allen, M., Gibbons, M., Jost, S., Kucaba, T.,  
Martin, J., Pape, D., Steptoe, M., Underwood, K., Thelning, B.,  
Ritter, E., Bowers, Y., Wylie, T., Waterston, R., and Wilson, R.

WashU Zebrafish EST Project 1999  
Unpublished (1999)

On Jul 7, 1999 this sequence version replaced gi:5866176.  
Contact: S.L. Johnson

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800

Email: est@wustl.edu  
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami  
Sequencing by: Washington University Genome Sequencing Center  
Seq primer: T7 ET from Amersham

High quality sequence stop: 55.  
Location/Qualifiers

## FEATURES



Seq primer: -28ml3 rev2 ET from Amer sham  
High quality sequence stop: 59.  
Location/Qualifiers

FEATURES  
source

```
1..79
/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/clone="IMAGE:578466"
/clone_lib="Barstead MPLR1"
/sex="mixed"
/tissue_type="Kidney"
/dev_stage="6 weeks"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCAGATCTGTAAGTGGAGCGCGCCGCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[CATGATTCGGTACC], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT7T3 vector.
Library constructed by Bob Barstead."
```

BASE COUNT  
ORIGIN

```
31 a 16 c 11 g 21 t
```

Query Match 59.0%; Score 11.8; DB 28; Length 79;  
Best Local Similarity 66.7%; Pred. No. 1.6e+04;  
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 6 rtangcngtctt 20  
| | | | | | | | | |  
Db 70 ATATGCGATTTT 56

RESULT 5  
LOCUS

A1749415/c 52 bp mRNA EST 22-JUN-1999  
DEFINITION at23g11.x1 Barstead aorta HPLRB6 Homo sapiens cDNA clone  
IMAGE:2356004 3' similar to SW:RL35\_HUMAN P42766 60S RIBOSOMAL  
PROTEIN L35. ; mRNA sequence.

ACCESSION A1749415 GI:5127679  
VERSION A1749415.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiinae; Homo.  
1 (bases 1 to 52)

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Gelsel,G., Jost,S.,  
Kritman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,  
Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,  
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
WashU-NCI human EST Project

JOURNAL Unpublished (1997)  
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189244.  
CONTACT: Wilson RK

WASHINGTON UNIVERSITY School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewartson.wustl.edu

This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Trace considered overall poor quality  
Possible reversed clone: similarity on wrong strand  
Seq primer: -40UP from Gibco  
High quality sequence stop: 1.  
Location/Qualifiers

FEATURES  
source

```
1..52
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2356004"
/clone_lib="Barstead aorta HPLRB6"
```

```
/sex="male"
/dev_stage="adult, age 64"
/lab_host="DH10B (phage resistant)"
/notes="Organ: aorta; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTCAGATCTGTAAGTGGAGCGCGCCGCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[5' AATCGGATCGAAC 3' and 5' GTTGATCGG 3'], digested
with Not I and cloned into the Not I and Eco RI sites of
the modified pT7T3 vector. Library constructed by Bob
Barstead."
```

BASE COUNT 15 a 11 c 13 g 13 t  
ORIGIN

Query Match 58.0%; Score 11.6; DB 51; Length 52;  
Best Local Similarity 57.9%; Pred. No. 2e+04;  
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 atncrtangcngtctt 19  
| | | | | | | | | |  
Db 33 AATCGATTCGCCGCTTTT 15

RESULT 6  
LOCUS

T99376 74 bp mRNA EST 31-MAR-1995  
DEFINITION Y664f04.s1 Soares fetal liver spleen INF1S Homo sapiens cDNA clone  
IMAGE:122527 3' similar to gb:S55700 HYDROXYMETHYLGLUTARYL-COA  
LYASE PRECURSOR (HUMAN); mRNA sequence.

ACCESSION T99376 GI:749113  
VERSION T99376.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiinae; Homo.  
1 (bases 1 to 74)

AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
Wilson,R.

JOURNAL The WashU-Merck EST Project  
COMMENT Unpublished (1995)  
CONTACT: Wilson RK

WASHINGTON UNIVERSITY School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewartson.wustl.edu

High quality sequence starts: 1 High quality sequence stops: 1  
Source: IMAGE Consortium, LNL This clone is available royalty-free  
through LNL; contact the IMAGE Consortium (info@image.lnl.gov)  
for further information. Trace considered overall poor quality  
Insert length: 1217 Std Error: 0.00  
Seq primer: -2ml13  
High quality sequence stop: 1.  
Location/Qualifiers

FEATURES  
source

```
1..74
/organism="Homo sapiens"
/db_xref="GDB:475072"
/db_xref="taxon:9606"
/clone="IMAGE:122527"
/clone_lib="Soares fetal liver spleen INF1S"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
```

BASE COUNT 22 a 13 c 17 g 20 t 2 others

Query Match	58.0%	Score 11.6	DB 21	Length 74
Best Local Similarity	61.1%	Pred. No. 2e+04		
Matches 11, Conservative	2, Mismatches	5, Indels	0, Gaps	0,
Oy	1 atnccrtangcngtftly	18		
	:			
Db	3 ATTACATAAGCTGINTTC	20		

[illegible]

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 93)	Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Mattlin, J., Moore, B., Schellenberg, K., Stepien, M., Tan, F., Thaising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.	WashU-NCI human EST project		Unpublished (1997)
				On Jun 5, 1998 this sequence version replaced gi:3188824.

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel.: 314 286 1800  
Fax: 314 286 1810  
Email: [estwatson.wustl.edu](mailto:estwatson.wustl.edu)  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
Seq primer: -40RP from gibco.

FEATURES	SOURCE
Location/Qualifiers	1. 93
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="IMAGE:2519540"	
/clone_lib="Schneider fetal brain 00004"	
/sex="male"	
/tissue_type="frontal lobe"	
/dev_stage="5 months post-conception"	
/lab_host="DHI0B"	
/note="Organ: brain; Vector: pBluescript SK (Stratagene); Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was prepared from human fetal brain tissue. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence:	
5'-GAGAGAGAGAGAGACTCGACAGATCCTTAATTAATTAATCCCCCCCCCC-3'	
and 3' adaptor sequence:	
5'-GAGAGAGAGAGAGACTCGAGTTTTTT-3'.	
The library was size-selected for >0.5 Kb inserts and has an average insert size estimated at 1.2 Kb. This library was constructed using the CAP-trapper method for full-length enrichment and has not undergone amplification. Library was constructed by Dr. Claudio Schneider (LNCIB-Area Science Park, Trieste, Italy)."	

## ORIGIN

Query Match	58.0%;	Score 11.6;	DB 62;	Length 93;
Best Local Similarity	64.7%;	Pred. No. 2e+04;		
Matches 11; Conservative	2;	Mismatches 4;	Indels 0;	Gaps 0;

```

QY      1 atnccrtangcnglttt 17
          ||||:|||||:|
Db      49 ATGCCATAAGCTGTGCT 33

```

LOCUS	94 bp	DNA	GSS	17-OCT-1997
DEFINITION	B35774			
	HS-1030-B1-E09-MR.abi	CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate-CT810 Col-17 Row-J, genomic survey sequence.		

ACCESSION	B35774	GI:2535143
VERSION	B35774.1	
KEYWORDS	GSS.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE  
1 (bases 1 to 94)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 94)		
Mahaitas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,...		
Tratcoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.		
Construction of a Characterized Clone Resource for Genomic		

Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors  
Unpublished (1997)

COMMENT

Contact: Mahairas GG, Zackrone KD, Hood  
University of Washington  
Seattle, WA 98195, USA  
Tel: (206) 616-8744  
Fax: (206) 685-7301  
Email: kzackrone@u.washington.edu  
Sequence Tagged Connector  
Plate: C7810 row: J column: 17  
Class: BAC ends  
High quality sequence stop: 94.

## FEATURES

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-CM810 Col=17 Row=J"
/clone_1b="Ctfr Human Genomic Sperm Library C"
/sex="M"
/note="Organ: Sperm; Vector: pBelDac11; BAC Clones in
E-coll DH10B"
BASE COUNT      18 a      20 c      13 g      41 t      2 others
ORIGIN

```

Query Match	58.0%;	Score 11.6;	DB 84;	Length 94;
Best Local Similarity	57.9%;	Pred. No. 2e+04;		
Matches 11; Conservative	3;	Mismatches 5;	Indels 0;	Gaps 0;

QY	1	atncrcrtangcngtrtlyt	19
		:      :	
Db	19	ATACTGTATCCTGTGTTT	37

RESULT	9			
AA449805	AA449805	95 bp	mRNA	EST
LOCUS	2209P01.s1	Soares	total_fetus_Nb2HF8_9w	04-JUN-1997
DEFINITION	IMAGE:786001.3	mRNA sequence.		
ACCESSION	AA449805			
VERSION	AA449805.1	GI:2163555		
KEYWORDS	EST.			
SOURCE	human.			

ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
AUTHORS	1 (bases 1 to 95) Hillier, L., Allen, M., Bowles, L., Dubuque, T., Giesel, G., Jost, S., Kuwab, T., Lacy, M., Le, N., Lennon, G., Marz, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyke, T., Waterston, R. and Wilson, R.			
TITLE	WashU-Merck EST Project 1997			
JOURNAL	Unpublished (1997)			
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1392947. Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -41m13 fwd. Et from Amersham High quality sequence stop: 94. Location/Qualifiers 1..95 /organism="Homo sapiens" /db_xref="GDB:5983520" /db_xref="taxon:9606" /clone="IMAGE:786001" /clone_1lb="Scars_total_fetus_NB2HF8_9w" /dev_stage="-8-9-weeks" /lab_host="DH10B" /note="Vector: pRTT3-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAAGTGGAGCGCGCTTAATTTTTTTTTTTT 3'] . Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRTT3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "			
BASE COUNT	26 a 16 c 14 g 39 t			
ORIGIN				
Query Match	58.0%; Score 11.6; DB 33; Length 95;			
Best Local Similarity	57.9%; Pred. No. 2e+04;			
Matches 11; Conservative	3; Mismatches 5; Indels 0; Gaps 0;			
Oy	2 tncctangcngtrtttytt 20   : :      : : :			
Db	9 TACCAATTGCTTATTTT 27			
RESULT 10				
LOCUS	D25841 58 bp mRNA EST 30-NOV-1995			
DEFINITION	HMGSO4213 Human colon mucosa Homo sapiens cDNA clone cml426 3', mRNA sequence.			
ACCESSION	D25841			
VERSION	D25841.1 GI:500509			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
AUTHORS	Okubo, K., Yoshii, J., Yokouchi, H., Kameyama, M. and Matsubara, K.			
TITLE	Global analysis of gene expression in colon mucosa: a large scale random cDNA sequencing analysis			
JOURNAL	Unpublished (1994)			
COMMENT	Contact: Okubo, K., Itob, K., Yoshii, J., Yokouchi, H. and Matsubara, K. Institute for Molecular and Cellular Biology Osaka University			

FEATURES	3-1 Yamada-Oka, Suita, Osaka 565, Japan.
SOURCE	Location/Qualifiers 1. .58
BASE COUNT	30 a 5 c 6 g 17 t
ORIGIN	
Query Match	57.0%; Score 11.4; DB 20; Length 58;
Best Local Similarity	60.0%; Pred. No. 2.6e+04; Indels 0; Gaps 0;
Matches 12; Conservative 2; Mismatches 6;	
Oy	1 atnccrtangcngtrtlytt 20    : :     : :
Db	26 ATACAGTATGAAGTTTCTT 7
RESULT 11	
LOCUS	H55643 80 bp DNA EST 07-DEC-1995
DEFINITION	CHR220582 Chromosome 22 exon Homo sapiens genomic clone C22_792 5', mRNA sequence.
ACCESSION	H55643
VERSION	H55643.1 GI:1108509
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 80) Trofater,J.A., Long,K.R., Murrell,J.R., Stotler,C.J., Gusella,J.F. and Buckler,A.J. An expression-independent catalog of genes from human chromosome 22 Genome Res. 5 (3), 214-224 (1995) 96159527 On May 18, 1995 this sequence version replaced g1.811655. Contact: Buckler AJ Molecular Neurogenetics Unit Massachusetts General Hospital Building 149, 13th St., Charlestown MA 02129 Tel: 6177249616 Fax: 6177265736 Email: buckler@helix.mgh.harvard.edu Insert Length: 1118 Std Error: 0.00 Seq primer: T3 High quality sequence stop: 319. Location/Qualifiers 1. .80 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="C22_792" /clone_1lb="Chromosome 22 exon" /lab_host="E. coli DH5a" /note="Vector: pBluescriptlTKS+. Site_1: Sal I; Site_2: Bam HI (destroyed); Exons were isolated from human chromosome 22 specific cosmids using a modification of the method of exon amplification (Proc. Natl. Acad. Sci. USA 88:4005-4009, 1991). Amplified exons were digested with Sal I and Bgl II and subsequently cloned into pBluescriptlTKS+ at the Sal I and Bam HI sites."
FEATURES	
SOURCE	
BASE COUNT	16 a 14 c 23 g 27 t
ORIGIN	
Query Match	57.0%; Score 11.4; DB 23; Length 80;
Best Local Similarity	60.0%; Pred. No. 2.6e+04; Indels 0; Gaps 0;
Matches 12; Conservative 2; Mismatches 6;	
Oy	1 atnccrtangcngtrtlytt 20



Db 56 AGGCCTAGGCTGCTTT 75

RESULT 12  
LOCUS AI318353  
DEFINITION ta78604.x1 NCI CGAP\_HSC2 Homo sapiens CDNA clone IMAGE:2050230 3'  
ACCESSION AI318353  
VERSION AI318353.1 GI:4034235  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 84)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Herbert Morse, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

FEATURES  
source  
1. .84  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2050230"  
/clone\_1lb="NCI-CGAP\_HSC2"  
/tissue\_type="stem cell 34+/38+"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: bone marrow; Vector: pAMP1; mRNA made from  
bone marrow, stem cells 34+/38+, CDNA made by oligo-dT  
priming. Directionally cloned. Size-selected on agarose  
gel, average insert size 400 bp. Primary library,  
non-amplified."

BASE COUNT 15 a 27 c 14 g 28 t  
ORIGIN

Query Match 57.0%; Score 11.4; DB 44; Length 84;  
Best Local Similarity 60.0%; Pred. NO. 2.6e+04;  
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 atnccrtangcngtrtytt 20  
|||:|||||:|||||  
Db 45 ATCCGATACGCGCTTCTT 64

RESULT 13  
LOCUS AI318677  
DEFINITION ta49e05.x1 NCI-CGAP\_Lu25 Homo sapiens CDNA clone IMAGE:2047424 3'  
ACCESSION AI318677  
VERSION AI318677.1 GI:4034612  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 87)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

FEATURES  
source  
1. .87  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2047424"  
/clone\_1lb="NCI-CGAP\_Lu25"  
/tissue\_type="bronchioalveolar carcinoma"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pAMP1; mRNA made from lung  
carcinoma tissue, CDNA made by oligo-dT priming.  
Directionally cloned. Size-selected on agarose gel,  
average insert size 500 bp. Primary library,  
non-amplified."

BASE COUNT 21 a 29 c 14 g 23 t  
ORIGIN

Query Match 57.0%; Score 11.4; DB 44; Length 87;  
Best Local Similarity 60.0%; Pred. NO. 2.6e+04;  
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 atnccrtangcngtrtytt 20  
|||:|||||:|||||  
Db 48 ATCCATACGCGCTTCTT 67

RESULT 14  
LOCUS AA910722  
DEFINITION ok4e12.s1 NCI-CGAP\_KiD3 Homo sapiens CDNA clone IMAGE:1520686 3'  
ACCESSION AA910722  
VERSION AA910722.1 GI:3050012  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 100)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 19, 1998 this sequence version replaced gi:2152941.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/dbrrp/image/image.html

Trace considered overall poor quality  
Insert Length: 370 Std Error: 0.00  
Seq primer: -40m3 fwd. ET from Amersham  
High quality sequence stop: 1.

## FEATURES

source location/Qualifiers

1..100  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1520686"  
/clone.lib="NCI CGAP\_K1d3"  
/lab\_host="DH10B"  
/note="organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer,  
(double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not  
I and Eco RI sites of the modified pT7T3 vector. mRNA  
source: 2 pooled kidneys. Library went through one round  
of normalization. Library constructed by Bento Soares and  
M. Fatima Bonaldo."

BASE COUNT 11 a 29 c 22 g 38 t

## ORIGIN

Query Match 57.0%; Score 11.4; DB 40; Length 100;  
Best local Similarity 60.0%; Pred. No. 2.6e+04;  
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 atnccrtangcngtrtlytt 20

1 |||:| |||||:|

Db 79 ACCCCATGAGCTGTTCTT 98

## RESULT 15

AI098711

LOCUS AI098711 100 bp mRNA EST 20-AUG-1998  
DEFINITION vg80b10.r1 Barstead MPLRB1 Mus musculus cDNA clone IMAGE:872251 5',  
similar to SW:COX1\_RAT P05503 CYTOCHROME C OXIDASE POLYPEPTIDE I ;,  
mRNA sequence.

ACCESSION

AI098711

VERSION AI098711.1 GI:3448236

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 100)

AUTHORS

Marta M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

TITLE

The WashU-HMI Mouse EST Project

JOURNAL

Unpublished (1996)

COMMENT

On Apr 14, 1993 this sequence version replaced gi:639245.

Contact: Marta M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:511731

Trace considered overall poor quality

Possible reversed clone: similarity on wrong strand

Seq primer: -28m3 rev2 ET from Amersham  
High quality sequence stop: 1.  
location/Qualifiers

## FEATURES

source

1..100  
/organism="Mus musculus"  
/strain="BALB/c"  
/db\_xref="taxon:10090"  
/clone="IMAGE:872251"  
/clone.lib="Barstead MPLRB1"  
/sex="mixed"  
/issue\_type="Kidney"  
/dev\_stage="6 weeks"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5',  
TGTTCGATCTGAGTGGAGCGGCCGCCCTTTTCTTTTCTTTTCTTTTCTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
[CATGGAATTCGTACC], digested with Not I and cloned into the  
Not I and Eco RI sites of the modified pT7T3 vector.  
Library constructed by Bob Barstead."

BASE COUNT 33 a 22 c 11 g 34 t

## ORIGIN

Query Match 57.0%; Score 11.4; DB 42; Length 100;  
Best local Similarity 60.0%; Pred. No. 2.6e+04;  
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 atnccrtangcngtrtlytt 20

1 |||:| |||||:|

Db 9 ATTTCGTACGCTCTATTGTT 28

Search completed: June 4, 2000, 13:53:05  
Job time: 20929 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:09:39 ; Search time 244.64 Seconds  
(Without alignments)  
10.627 Million cell updates/sec

Title: US-09-164-714-5

Perfect score: 20  
Sequence: 1 atnccrtangcngtrtlytt 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 375880

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/PCRTUS.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12.6	63.0	97	3	US-08-852-807-17
2	12.2	61.0	48	1	US-08-060-822A-18
3	12.2	61.0	48	6	PCT-US94-05257-18
4	12.2	61.0	77	1	US-08-399-412A-18
5	12	60.0	49	3	US-07-673-661B-2
6	11.8	59.0	61	4	US-08-483-528B-50
7	11.8	59.0	61	5	US-08-673-799C-50
8	11.8	59.0	87	1	US-08-425-336-135
9	11.8	59.0	87	2	US-08-488-113B-135
10	11.8	59.0	87	2	US-08-477-484B-135
11	11.8	59.0	87	2	US-08-107-669D-60
12	11.8	59.0	87	2	US-08-477-788A-60
13	11.8	59.0	87	2	US-08-477-531B-60
14	11.8	59.0	87	3	US-08-646-360-135
15	11.8	59.0	87	3	US-08-082-842A-60
16	11.6	58.0	44	1	US-07-603-451A-10
17	11.6	58.0	44	1	US-08-060-822A-11
18	11.6	58.0	44	6	PCT-US94-05257-11
19	11.6	58.0	48	1	US-08-060-822A-19
20	11.6	58.0	48	6	PCT-US94-05257-19
21	11.4	57.0	29	1	US-08-233-788A-8
22	11.4	57.0	34	4	US-08-924-695A-17
23	11.4	57.0	40	3	US-08-281-423-7
24	11.2	56.0	39	3	US-08-470-419-6
25	11.2	56.0	39	3	US-08-761-828-6
26	11.2	56.0	39	4	US-08-290-105-6
27	11.2	56.0	39	5	US-08-776-949-6

#### ALIGNMENTS

28	11.2	56.0	48	3	US-08-470-419-20	Sequence 20, Appl
29	11.2	56.0	48	3	US-08-761-828-20	Sequence 20, Appl
30	11.2	56.0	48	4	US-08-290-105-20	Sequence 20, Appl
31	11.2	56.0	48	5	US-08-776-949-20	Sequence 20, Appl
32	11.2	56.0	62	6	PCT-US96-09455A-378	Sequence 378, Appl
33	11	55.0	27	1	US-08-399-412A-97	Sequence 97, Appl
34	11	55.0	29	2	US-08-623-891-107	Sequence 107, App
35	11	55.0	63	1	US-07-901-708-3	Sequence 3, Appl
36	10.6	53.0	22	3	US-08-117-952-111	Sequence 11, App
37	10.6	53.0	29	1	US-08-495-743-26	Sequence 26, Appl
38	10.6	53.0	29	2	US-08-495-739-26	Sequence 26, Appl
39	10.6	53.0	29	2	US-08-495-741-26	Sequence 26, Appl
40	10.6	53.0	30	3	US-08-583-276-5	Sequence 5, Appl
41	10.6	53.0	31	2	US-08-403-762A-70	Sequence 70, Appl
42	10.6	53.0	31	3	US-08-829-961-4	Sequence 4, Appl
43	10.6	53.0	34	5	US-08-886-466-8	Sequence 8, Appl
44	10.6	53.0	40	5	US-08-841-267-4	Sequence 4, Appl
45	10.6	53.0	77	1	US-08-399-412A-14	Sequence 14, Appl

RESULT 1  
US-08-852-807-17/C  
; Sequence 17, Application US/08852807  
; Patent No. 5861298  
; GENERAL INFORMATION:  
; APPLICANT: Debouck, Christine  
; APPLICANT: Drake, Fred  
; APPLICANT: Gowen, Maxine  
; APPLICANT: Hood, Julie  
; APPLICANT: Hastings, Gregg  
; APPLICANT: Adams, Mark  
; APPLICANT: Fraser, Claire  
; APPLICANT: Lee, No. 5861298man  
; APPLICANT: Kirness, Even  
; APPLICANT: Blake, Judith  
; APPLICANT: Fitzgerald, Lisa  
; TITLE OF INVENTION: CATHEPSIN K GENE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-2799  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA: 807  
; APPLICATION NUMBER: US/08/852, 807  
; FILING DATE: 07-MAY-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/019, 942  
; FILING DATE: 14-JUNE-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/020, 273  
; FILING DATE: 17-JUNE-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/026, 273  
; FILING DATE: 26-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Han, William T  
; REGISTRATION NUMBER: 34,344  
; REFERENCE/DOCKET NUMBER: ATG50006-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-5219  
; TELEFAX: 610-270-5090

TELEX:  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 97 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE:  
; ORIGINAL SOURCE:  
US-08-852-807-17

Query Match 63.0%; Score 12.6; DB 3; Length 97;  
Best local Similarity 60.0%; Pred. No. 1.4e+02;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 atnccrtangngtrtlytt 20  
|||: ||| |||: |||  
Db 69 ATCCACAGCGCTGTCTT 50

RESULT 2  
US-08-060-822A-18/c  
; Sequence 18, Application US/08060822A  
; Patent No. 5432270  
; GENERAL INFORMATION:  
; APPLICANT: Bevins, Charles L.  
; APPLICANT: Diamond, Gill  
; APPLICANT: Zaslloff, Michael  
; TITLE OF INVENTION: No. 5432270e1 Antimicrobial Peptides  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5432270ris  
; STREET: One Liberty Place - 46th Floor  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: WORDPERECT 5.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/060,822A  
; FILING DATE: 19930511  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/991,200  
; FILING DATE: December 15, 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rebecca Ralph Gaumond  
; REGISTRATION NUMBER: 35,152  
; REFERENCE/DOCKET NUMBER: CH-0408  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-3100  
; TELEFAX: (215) 568-3439  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 48 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-060-822A-18

Query Match 61.0%; Score 12.2; DB 1; Length 48;  
Best local Similarity 61.1%; Pred. No. 2.1e+02;  
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 atnccrtangngtrtlytt 18  
|||: ||| |||: |||  
Db 47 ATCCTGTAACTGTGT 30

RESULT 3  
PCT-US94-05257-18/c  
; Sequence 18, Application PC/TUS9405257  
; GENERAL INFORMATION:  
; APPLICANT: Bevins, Charles L.  
; APPLICANT: Diamond, Gill  
; APPLICANT: Zaslloff, Michael  
; TITLE OF INVENTION: Novel Antimicrobial Peptides  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris  
; STREET: One Liberty Place - 46th Floor  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: WORDPERECT 5.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/05257  
; FILING DATE: 11-MAY-1994  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/060,822  
; FILING DATE: May 11, 1993  
; APPLICATION NUMBER: 07/991,200  
; FILING DATE: December 15, 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rebecca Ralph Gaumond  
; REGISTRATION NUMBER: 35,152  
; REFERENCE/DOCKET NUMBER: CH-0408  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-3100  
; TELEFAX: (215) 568-3439  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 48 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
PCT-US94-05257-18

Query Match 61.0%; Score 12.2; DB 6; Length 48;  
Best local Similarity 61.1%; Pred. No. 2.1e+02;  
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 atnccrtangngtrtlytt 18  
|||: ||| |||: |||  
Db 47 ATCCTGTAACTGTGT 30

RESULT 4  
US-08-399-412A-18  
; Sequence 18, Application US/08399412A  
; Patent No. 5622828  
; GENERAL INFORMATION:  
; APPLICANT: Patma, David  
; APPLICANT: Gold, Larry  
; TITLE OF INVENTION: High-Affinity Oligonucleotide  
; TITLE OF INVENTION: Ligands To Secretory Phospholipase  
; NUMBER OF SEQUENCES: 122  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Swanson & Bratschun, L.L.C.



TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "synthetic DNA"  
US-08-483-528B-50

Query Match 59.0%; Score 11.8; DB 4; Length 61;  
Best Local Similarity 66.7%; Pred. No. 3.6e+02;  
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 6 rtangcngrtlytt 20  
Db 26 GTAGCGTGTCTTCTT 12

RESULT 7  
US-08-673-799C-50/c  
Sequence 50, Application US/08673799C  
Patent No. 6042828

GENERAL INFORMATION:  
APPLICANT: NAKAMURA, KAZUYASU  
APPLICANT: KOIKE, MASAMICHI  
APPLICANT: SHITARA, KENYA  
APPLICANT: HANAI, NOBUO  
APPLICANT: KURAMA, YOSHIOHISA  
APPLICANT: HASEGAWA, MAMORU  
TITLE OF INVENTION: HUMANIZED ANTIBODIES  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHVE P.C.  
STREET: 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/673,799C  
FILING DATE: 27-JUN-96  
CLASSIFICATION: 536  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)816-4000  
TELEFAX: (703)816-4100

INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 61 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "synthetic DNA"

US-08-673-799C-50

Query Match 59.0%; Score 11.8; DB 5; Length 61;  
Best Local Similarity 66.7%; Pred. No. 3.6e+02;  
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 6 rtangcngrtlytt 20  
Db 26 GTAGCGTGTCTTCTT 12

RESULT 8  
US-08-425-336-135/c  
Sequence 135, Application US/08425336  
Patent No. 5621083  
GENERAL INFORMATION:  
APPLICANT: Belter, Marc D.

APPLICANT: Carroll, Stephen F.  
APPLICANT: Studnika, Gary M.  
TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating  
TITLE OF INVENTION: Proteins  
NUMBER OF SEQUENCES: 140  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA

ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/425,336  
FILING DATE: 18-APR-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/064,691  
FILING DATE: 12-MAY-1993  
APPLICATION NUMBER: US 07/901,707  
FILING DATE: 19-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/787,567  
FILING DATE: 04-NOV-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Thomas C.  
REGISTRATION NUMBER: P-36,989  
REFERENCE/DOCKET NUMBER: 31394  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 135:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 87 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA

US-08-425-336-135

Query Match 59.0%; Score 11.8; DB 1; Length 87;  
Best Local Similarity 66.7%; Pred. No. 3.8e+02;  
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 6 rtangcngrtlytt 20  
Db 44 ATAGGAGTGTCTTCTT 30

RESULT 9  
US-08-488-113B-135/c  
Sequence 135, Application US/08488113B  
Patent No. 5744380

GENERAL INFORMATION:  
APPLICANT: Belter, Marc D.  
APPLICANT: Carroll, Stephen F.  
APPLICANT: Studnika, Gary M.  
TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating  
TITLE OF INVENTION: Proteins  
NUMBER OF SEQUENCES: 169  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McAndrews, Held & Malloy, Ltd.  
STREET: 500 West Madison Street, 34th floor  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA



```

;
; ZIP: 60661
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,113B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/425,336
; FILING DATE: 18-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,691
; FILING DATE: 12-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/988,430
; FILING DATE: 09-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/901,707
; FILING DATE: 19-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/787,567
; FILING DATE: 04-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McNicholas, Janet M.
; REGISTRATION NUMBER: 32,918
; REFERENCE/DOCKET NUMBER: 11022US07/200-70.P3.C2A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/707-8889
; TELEFAX: 312/707-9155
; TELEX: 650 388-1248
; INFORMATION FOR SEQ ID NO: 135:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 87 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-488-113B-135

Query Match          59.0%; Score 11.8; DB 2; Length 87;
Best Local Similarity 66.7%; Pred. No. 3.8e+02;
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy      6 rtangcngtrtlytt 20
      :|||1111111111
Db      44 ATAGGAGGTGTCTT 30

RESULT 10
US-08-477-484B-135/C
; Sequence 135, Application US/08477484B
; Patent No. 5756699
; GENERAL INFORMATION:
; APPLICANT: Better, Marc D.
; APPLICANT: Carroll, Stephen F.
; APPLICANT: Studnicka, Gary M.
; TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating
; PROTEINS
; NUMBER OF SEQUENCES: 169
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: McAndrews, Held & Malloy, Ltd.
; STREET: 500 West Madison Street, 34th floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60661
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,484B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/425,336
; FILING DATE: 18-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,691
; FILING DATE: 12-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/988,430
; FILING DATE: 09-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/901,707
; FILING DATE: 19-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/787,567
; FILING DATE: 04-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McNicholas, Janet M.
; REGISTRATION NUMBER: 32,918
; REFERENCE/DOCKET NUMBER: 11022US07/200-70.P3.C2A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/707-8889
; TELEFAX: 312/707-9155
; TELEX: 650 388-1248
; INFORMATION FOR SEQ ID NO: 135:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 87 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-477-484B-135

Query Match          59.0%; Score 11.8; DB 2; Length 87;
Best Local Similarity 66.7%; Pred. No. 3.8e+02;
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy      6 rtangcngtrtlytt 20
      :|||1111111111
Db      44 ATAGGAGGTGTCTT 30

RESULT 11
US-08-107-669D-60/C
; Sequence 60, Application US/08107669D
; Patent No. 5766886
; GENERAL INFORMATION:
; APPLICANT: Studnicka, Gary M.
; TITLE OF INVENTION: Modified Antibody Variable Domains (as amended)
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sterne, Kessler, Goldstein and Fox P.L.L.C.
; STREET: 1100 New York Ave., N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/107,669D
; FILING DATE: 13-AUG-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
```

```

1  APPLICATION NUMBER:  PCT/US92/10906
2  FILING DATE:  14-DEC-1992
3  PRIOR APPLICATION DATA:
4  APPLICATION NUMBER:  US 07/808,464
5  FILING DATE:  13-DEC-1991
6  ATTORNEY/AGENT INFORMATION:
7  NAME:  Michele A. Cimdala
8  REGISTRATION NUMBER:  33,851
9  REFERENCE/DOCKET NUMBER:  0610.1000001/MAC
10 TELECOMMUNICATION INFORMATION:
11 TELEPHONE:  202/371-2600
12 TELEFAX:  202/371-2540
13 INFORMATION FOR SEQ ID NO:  60:
14 SEQUENCE CHARACTERISTICS:
15     LENGTH:  87 base pairs
16     TYPE:  nucleic acid
17     STRANDEDNESS:  single
18     TOPOLOGY:  linear
19     MOLECULE TYPE:  DNA
20     US-08-107-669D-60

```

Query Match	59.0%;	Score 11.8;	DB 2;	Length 87;
Best Local Similarity	66.7%;	Pred. No. 3.8e+02;		
Matches 10;	Conservative 3;	Mismatches 2;	Indels 0;	Gaps 0

Qy	6	rtangcngtrtlyt	20
	:	:   :	
Db	44	ATAGGCAGTGTTC	30

RESULT 12  
 US-08-472-788A-60/C  
 Sequence 60, Application US/08472788A  
 Patent No. 5770196  
 GENERAL INFORMATION:  
 APPLICANT: Studnicka, Gary M.  
 TITLE OF INVENTION: Modified Antibody Variable Domains  
 NUMBER OF SEQUENCES: 89  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Stearns, Kessler, Goldstein and Fox P.L.L.C.  
 STREET: 1100 New York Ave., N.W., Suite 600  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20005-3934  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/472,788A  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/082,842  
 FILING DATE: 23-JUN-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US92/10906  
 FILING DATE: 14-DEC-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/808,464  
 FILING DATE: 13-DEC-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Cimbala, Michele A.  
 REGISTRATION NUMBER: 33,851  
 REFERENCE/DOCKET NUMBER: 0610.1000003  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202/371-2600  
 TELEFAX: 202/371-2540  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 60:

```

; SEQUENCE CHARACTERISTICS
; LENGTH: 87 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-472-788A-60

```

Query Match	59.0%;	Score 11.8;	DB 2;	Length 87;
Best Local Similarity	66.7%;	Pred. No. 3.8e+02;		
Matches 10;	Conservative 3;	Mismatches 2;	Indels 0;	Gaps 0;

```
QY      6 rtangcngtrtlyt 20
          :|| || ||:||:|
Db      44 ATAGGCAGTGTTCTT 30
```

```

RESULT 13
US-08-477-531B-60/C
; Sequence 60, Application US/08477531B
; Patent No. 3621123
;
; GENERAL INFORMATION:
; APPLICANT: Studicka, Gary M.
; TITLE OF INVENTION: Modified Antibody Variable Domains (as amended)
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein and Fox P.L.L.C.
; STREET: 1100 New York Ave., N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20005-3934
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,531B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 436
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/107,669
; FILING DATE: 13-AUG-1993
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/10906
; FILING DATE: 14-DEC-1992
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/808,464
; FILING DATE: 13-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Michele A. Cimbalia
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 0610.1000004/MAC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/371-2540
; TELEFAX: 202/371-2540
;
; INFORMATION FOR SEQ ID NO: 60:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 87 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
;
; US-08-477-531B-60

```

Qy 6 rtangcngtrtlytt 20

Query Match	59.0%	Score 11.8;	DB 3	Length 87;
Best Local Similarity	66.7%	Pred. No. 3.8e02		
Matches	10;	Conservative	3;	Mismatches 2;
				Indels 0;
				Gaps 0.
QY	6	rtangcngtrttt	ttc	20
		. . .	. .	
Db	44	ATAGGCAAGTTCCTT	30	

Search completed: June 4, 2000, 16:09:40  
Job time: 28059 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:05:30 ; Search time 1236.38 Seconds

(without alignments)  
-22.031 Million cell updates/sec

Title: US-09-164-714-8

Perfect score: 28  
Sequence: 1 cccatgtgtgtcccaaatgtgtcaag 28

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 356616

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pr1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_v1:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_or:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_v1:\*  
32: gb\_htg1:\*  
33: gb\_htg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pr4:\*  
41: gb\_htg3:\*  
42: gb\_htg4:\*  
43: gb\_htg5:\*  
44: gb\_htg6:\*

45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pr5:\*  
52: gb\_htg8:\*  
53: gb\_htg9:\*  
54: gb\_htg10:\*  
55: gb\_htg11:\*  
56: gb\_htg12:\*  
57: gb\_htg13:\*  
58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	16	57.1	39	5 AR012248	AR012248 Sequence
C 2	16	57.1	39	5 I14946	I14946 Sequence 32
C 3	16	57.1	39	5 I73666	I73666 Sequence 32
4	15	53.6	34	5 A39689	A39689 Sequence 2
5	15	53.6	34	5 A48961	A48961 Sequence 8
6	14.8	52.9	75	40 AF064903	AF064903 Homo sapi
7	14.8	52.9	92	10 HUMPHOS15	MS9733 Human muscl
8	14.2	50.7	53	4 FR08001	AU008001 Fugu rudr
9	14.2	50.7	60	8 AF101150	AF101150 Brassica
C 10	14.2	50.7	72	5 AR040932	AR040932 Sequence
C 11	14.2	50.7	72	5 AR051839	AR051839 Sequence
C 12	14.2	50.7	72	5 I36229	I36229 Sequence 31
C 13	14.2	50.7	90	9 GORINSORFA	M69096 Gorilla end
C 14	14.2	50.7	99	34 DDIDKE	MS9748 D. discoiden
15	14	50.0	27	5 E14994	E14994 PCR primer
16	14	50.0	36	5 A78745	A78745 Sequence 6
17	14	50.0	36	5 AR014718	AR014718 Sequence
18	13.8	49.3	35	5 AR055776	AR055776 Sequence
C 19	13.8	49.3	35	5 AR055777	AR055777 Sequence
20	13.8	49.3	51	12 U92178	U92178 Mus musculu
C 21	13.6	48.6	26	5 AR014478	AR014478 Sequence
C 22	13.6	48.6	59	10 S59972	S59972 TAL1d1-prot
C 23	13.6	48.6	60	10 S72312	S72312 phosphofruc
24	13.6	48.6	81	3 PIGCROG	M25036 Hog prognat
25	13.6	48.6	99	12 MUSLRP03	L13679 Mouse leuko
26	13.4	47.9	44	5 AR048165	AR048165 Sequence
27	13.4	47.9	77	13 R1CL254A	D25437 Rice genom
C 28	13.4	47.9	87	10 S71607	S71607 pol clone
C 29	13.2	47.1	23	5 AR063157	AR063157 Sequence
C 30	13.2	47.1	29	5 AR003150	AR003150 Sequence
C 31	13.2	47.1	29	5 AR003243	AR003243 Sequence
C 32	13.2	47.1	29	5 AR009108	AR009108 Sequence
C 33	13.2	47.1	29	5 AR011395	AR011395 Sequence
C 34	13.2	47.1	29	5 AR028081	AR028081 Sequence
C 35	13.2	47.1	29	5 AR052797	AR052797 Sequence
C 36	13.2	47.1	29	5 I18033	I18033 Sequence 26
37	13.2	47.1	43	5 A05133	A05133 Oligonucleo
38	13.2	47.1	50	9 HPHDWIN2	K01660 Hepatitlis B
C 39	13.2	47.1	72	5 I00803	I00803 Sequence 7
C 40	13.2	47.1	77	5 A00281	A00281 Modified fr
41	13.2	47.1	77	5 A16012	A16012 HSA oligonu
C 42	13.2	47.1	93	12 MMU66712	U66712 Mus musculu
C 43	13	46.4	21	5 AR024394	AR024394 Sequence
C 44	13	46.4	21	5 I71186	I71186 Sequence 38
C 45	13	46.4	21	5 I83745	I83745 Sequence 38

#### ALIGNMENTS

[illegible]

FEATURES	source	1.	39	/organism="unknown"	Location/Qualifiers
BASE COUNT	9 a	16 c	8 g	6 t	
ORIGIN					
Query Match	57.18;	Score 16;	DB 5;	Length 39;	
Best Local Similarity	79.28;	Pred. No. 2.5e+03;			
Matches 19;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;	
Oy	2	cctatgtgtgtgccaaattgttc	25		
Db	34	CCAAATGTGTGTGGCCAGAAATGTGTC	11		
RESULT	4				
LOCUS	A39689	34 bp	DNA	PAT	05-MAR-1997
DEFINITION	Sequence 2 from Patent WO9418327.				
ACCESSION	A39689				
VERSION	A39689.1	GI:2295950			
KEYWORDS					
SOURCE	unidentified.				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 34)				
AUTHORS	Burnham, M.K., Chopra, I., Critchley, I.A. and Knowles, D.J.				
TITLE	FIBRONECTIN BINDING PROTEIN; MONOCLONAL ANTIBODY AND THEIR USE IN PREVENTING BACTERIAL ADHESION				
JOURNAL	Patent: WO 9418327-A 2 18-AUG-1994;				
COMMENT	SMITHKLINE BEECHAM PLC (GB)				
FEATURES	Other publication AU 5975994 940829				
source	Other publication CN 1119026 960320				
	Other publication CA 2155413 940818				
	Other publication ZA 9400728 941107				
	Other publication JP 8506810P 960723.				
	Location/Qualifiers				
	1.	34			
BASE COUNT	5 a	6 c	14 g	9 t	
ORIGIN	/organism="unidentified"				
	/db_xref="taxon:32644"				
Query Match	53.68;	Score 15;	DB 5;	Length 34;	
Best Local Similarity	78.38;	Pred. No. 7.2e+03;			
Matches 18;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;	
Oy	1	cctatgtgtgtgccaaattgttc	23		
Db	7	CCTTACGTTGTGTGGCAGCATTGG	29		
RESULT	5				
LOCUS	A48961	34 bp	DNA	PAT	07-MAR-1997
DEFINITION	Sequence 8 from Patent WO9604380.				
ACCESSION	A48961				
VERSION	A48961.1	GI:2302599			
KEYWORDS					
SOURCE	unidentified.				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 34)				
AUTHORS	Critchley, I.A., Dodd, I., Barnett, P. and Mossakowska, D.E.				
TITLE	D203 POLYPEPTIDE FROM STAPHYLOCOCCUS AUREUS AND USES THEREOF				
JOURNAL	Patent: WO 9604380-A 8 15-FEB-1996;				
COMMENT	SMITHKLINE BEECHAM PLC (GB)				
FEATURES	Other publication AU 3223895 960304.				
	Location/Qualifiers				

```
source 1..34
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 5 a 6 c 14 g 9 t
ORIGIN

Query Match 53.6%; Score 15; DB 5; Length 34;
Best Local Similarity 78.3%; Pred. No. 7.2e+03;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 cccatgtgtgccaaattgtcag 23
||| ||||| ||| |||||
Db 7 CCTTACTGTGGTGCACGATTGG 29

RESULT 6
AF064903 75 bp mRNA PRI 03-MAR-1999
LOCUS Homo sapiens clone 12-1 immunoglobulin heavy chain alpha VDJ region
DEFINITION mRNA, partial cds.
ACCESSION AF064903
VERSION AF064903.1 GI:4321896
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 75)
AUTHORS Rowley,A.H., Eckerley,C.A., Shulman,S.T. and Baker,S.C.
TITLE Evidence of an antigen-driven Iga immune response in the vascular
wall in acute Kawasaki Syndrome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 75)
AUTHORS Rowley,A.H., Eckerley,C.A., Shulman,S.T. and Baker,S.C.
TITLE Direct Submission
JOURNAL Submitted (14-MAR-1998) Pediatrics, Loyola University, 2160 S First
Avenue, Maywood, IL 60153, USA
FEATURES
source
1..75
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="12-1"
/tissue_type="vascular"
/note="from aorto-iliac junction of a 10-year-old child
with fatal acute Kawasaki Syndrome"
<1..>75
/note="includes IGHV3 gene"
/codon_start=1
/product="immunoglobulin heavy chain alpha VDJ region"
/protein_id="AADI5867.1"
/db_xref="GI:4321897"
/translation="AVYYCAKNQOVTPPYHYVMDVWG"

BASE COUNT 19 a 22 c 15 g 19 t
ORIGIN

Query Match 52.9%; Score 14.8; DB 40; Length 75;
Best Local Similarity 73.1%; Pred. No. 9.4e+03;
Matches 19; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 3 ctatgtgtgtccaaattgtcag 28
||||| ||||| ||| |||||
Db 6 CTATTACTGTGCGAATAATGCTCAAG 31

RESULT 7
HUMMPHOSIS 92 bp DNA PRI 05-DEC-1995
LOCUS Human muscle-type phosphofructokinase (PFK-M) gene, exon 1b.
DEFINITION M59733
ACCESSION M59733.1 GI:188647
VERSION
KEYWORDS muscle-type phosphofructokinase.
```

```
SEGMENT 15 of 23
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 92)
AUTHORS Yamasaki,T., Nakajima,H., Kono,N., Hotta,K., Yamada,K., Imai,E.,
Kuwajima,M., Noguchi,T., Tanaka,T. and Tani,S.
TITLE Structure of the entire human muscle phosphofructokinase-encoding
JOURNAL Gene 104 (2), 277-282 (1991)
MEDLINE 92009225
FEATURES
source
1..92
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="1cen-q32"
16..86
/feature="PFK-M"
/note="G00-120-277"
/number=16

BASE COUNT 23 a 17 c 32 g 20 t
ORIGIN

Query Match 52.9%; Score 14.8; DB 10; Length 92;
Best Local Similarity 73.1%; Pred. No. 9.5e+03;
Matches 19; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 3 ctatgtgtgtccaaattgtcag 28
||||| ||||| ||| |||||
Db 36 CTATTGTGGGGCTGGACTGGCCCAAG 61

RESULT 8
FRU8001/C
LOCUS FRU8001 53 bp DNA VRT 18-JAN-1999
DEFINITION Fugu rubripes Eya4 gene, exon 4.
ACCESSION AJ008001
VERSION AJ008001.1 GI:4164315
KEYWORDS EYA4 gene.
SOURCE Fugu rubripes.
ORGANISM Fugu rubripes.
Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
Tetraodontiformes; Tetraodontoidel; Tetraodontidae; Fugu.
REFERENCE 1 (bases 1 to 53)
AUTHORS Hanson,I.M.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-1998) Hanson I.M., Cell Genetics Section, MRC
Human Genetics Unit, Western General Hospital, Crewe Road,
Edinburgh, EH4 2XU, UK
2 (bases 1 to 53)
REFERENCE Borsani,G., deGrandi,A., Ballabio,A., Bulfone,A., Bernard,L.,
Banfi,S., Gattuso,C., Mariani,M., Dixon,M., Donna,D., Metcalfe,K.,
Winter,R., Robertson,M., Axton,R., Brown,A., van Heyningen,V. and
Hanson,I.
TITLE EYA4, a novel vertebrate gene related to dirosophila eyes absent
JOURNAL Hum. Mol. Genet. 8 (1), 11-23 (1999)
MEDLINE 99105912
FEATURES
source
1..53
Location/Qualifiers
/organism="Fugu rubripes"
/db_xref="taxon:31033"
/clone="cosmid 032C09"
1..53
/feature="Eya4"
/number=4
/usedin="AJ007998:Eya4_CDS"

BASE COUNT 14 a 11 c 12 g 16 t
ORIGIN
```

Query Match 50.7%; Score 14.2; DB 4; Length 53;  
Best Local Similarity 84.2%; Pred. No. 1.8e+04;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 atgttggtgccaattgg 23  
||||| ||||| |||

Db 49 ATGTGTGTGCAAAACGG 31

RESULT 9  
AF101150 60 bp DNA PLN 29-MAR-1999  
LOCUS Brassica napus clone na8 SINE S1 integration site sequence.  
ACCESSION AF101150  
VERSION AF101150.1 GI:4530252  
KEYWORDS  
SOURCE rape.  
ORGANISM Brassica napus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
eudicots; Rosidae; euroids II; Brassicales; Brassicaceae;  
Brassica.

REFERENCE 1 (bases 1 to 60)  
AUTHORS Tatout,C., Lavie,L. and Deragon,J.M.  
TITLE Similar target site selection occurs in integration of plant and  
mammalian retroposons  
JOURNAL J. Mol. Evol. (1999) In press  
REFERENCE 2 (bases 1 to 60)  
AUTHORS Tatout,C., Lavie,L. and Deragon,J.M.  
TITLE Direct Submision  
JOURNAL Submitted (26-OCT-1998) UMR 6547 BIOMOVE, CNRS, Universite Blaise  
Pascal, 24 Av. des landais, Aubiere 63177, France

FEATURES  
source  
1..60  
/organism="Brassica napus"  
/db\_xref="taxon:3708"  
/clone="na8"  
28..35  
/note="position of the direct repeats formed after S1  
integration"

misc\_feature  
integration"

BASE COUNT 19 a 6 c 15 g 20 t

ORIGIN

Query Match 50.7%; Score 14.2; DB 8; Length 60;  
Best Local Similarity 84.2%; Pred. No. 1.8e+04;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 atgttggtgccaattgg 23  
||||| ||||| |||||

Db 31 ATGTGTGTGCAAAATTCG 49

RESULT 10  
AR040932 72 bp DNA PAT 29-SEP-1999  
LOCUS AR040932  
DEFINITION Sequence 31 from patent US 5811238.  
ACCESSION AR040932  
VERSION AR040932.1 GI:5961428  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
UNCLASSIFIED.  
REFERENCE 1 (bases 1 to 72)  
AUTHORS Stemmer,W.P.C. and Cramer,I.A.  
TITLE Methods for generating polynucleotides having desired  
characteristics by iterative selection and recombination  
JOURNAL Patent: US 5811238-A 31 22-SEP-1998;  
FEATURES  
source  
1..72  
/organism="unknown"

BASE COUNT 21 a 17 c 18 g 16 t

ORIGIN

Query Match 50.7%; Score 14.2; DB 5; Length 72;  
Best Local Similarity 70.4%; Pred. No. 1.8e+04;  
Matches 19; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 ccatgttggtgccaattgcaag 28  
||||| ||||| |||||

Db 65 CTTCGTGTGACCAATATATGTAACG 39

RESULT 11  
AR051839 72 bp DNA PAT 29-SEP-1999  
LOCUS AR051839/c  
DEFINITION Sequence 31 from patent US 5830721.  
ACCESSION AR051839  
VERSION AR051839.1 GI:5975203  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
UNCLASSIFIED.  
REFERENCE 1 (bases 1 to 72)  
AUTHORS Stemmer,W.P.C. and Cramer,I.A.  
TITLE DNA mutagenesis by random fragmentation and reassembly  
JOURNAL Patent: US 5830721-A 31 03-NOV-1998;  
FEATURES  
source  
1..72  
/organism="unknown"

BASE COUNT 21 a 17 c 18 g 16 t

ORIGIN

Query Match 50.7%; Score 14.2; DB 5; Length 72;  
Best Local Similarity 70.4%; Pred. No. 1.8e+04;  
Matches 19; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 ccatgttggtgccaattgcaag 28  
||||| ||||| |||||

Db 65 CTTCGTGTGACCAATATATGTAACG 39

RESULT 12  
I36229 72 bp DNA PAT 13-MAY-1997  
LOCUS I36229/c  
DEFINITION Sequence 31 from patent US 5605793.  
ACCESSION I36229  
VERSION I36229.1 GI:2086742  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
UNCLASSIFIED.  
REFERENCE 1 (bases 1 to 72)  
AUTHORS Stemmer,W.P.C.  
TITLE Methods for in vitro recombination  
JOURNAL Patent: US 5605793-A 31 25-FEB-1997;  
FEATURES  
source  
1..72  
/organism="unknown"

BASE COUNT 21 a 17 c 18 g 16 t

ORIGIN

Query Match 50.7%; Score 14.2; DB 5; Length 72;  
Best Local Similarity 70.4%; Pred. No. 1.8e+04;  
Matches 19; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 ccatgttggtgccaattgcaag 28  
||||| ||||| |||||

Db 65 CTTCGTGTGACCAATATATGTAACG 39



Best Local Similarity	70.4%;	Pred. No. 1.8e+04;	
Matches 19;	Conservative 0;	Mismatches 8;	Indels 0;
Gaps	0;		

[illegible]

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:24:04 ; Search time 322.35 Seconds  
(without alignments)  
21.732 Million cell updates/sec

Title: US-09-164-714-8

Perfect score: 28  
Sequence: 1 cccatgtgtgtgcacaaatgtgtcaag 28

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 431286

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	15.8	56.4	24	1	X22061	PCR primer for C.
C 2	15.2	54.3	29	1	V43049	5' PCR primer Deg-
C 3	15	53.6	34	1	071835	Primer F1B2. Fibro
C 4	15	53.6	34	1	T18304	Primer for amplify
C 5	14.4	51.4	54	1	X27477	Oligo 1131 for hCG
C 6	14.4	51.4	60	1	X27467	Oligo 850 for hCG
C 7	14.4	51.4	84	1	X27464	Oligo 839 for hCG
C 8	14.2	50.7	72	1	Q97403	Mutated complement
C 9	14.2	50.7	72	1	T73061	Mutant light chain
C 10	14.2	50.7	72	1	V54920	Mutant light chain
C 11	14	50.0	27	1	V06529	Equine chorionic g
C 12	13.8	49.3	35	1	Q52717	Sequence of PCR pr
C 13	13.8	49.3	35	1	Q52718	Sequence of PCR pr
C 14	13.8	49.3	67	1	T24754	Human gene signatu
C 15	13.8	49.3	71	1	T65550	Staphylococcus aur
C 16	13.6	48.6	21	1	V37051	PCR primer for ant
C 17	13.6	48.6	41	1	X19306	Human granulocyte
C 18	13.6	48.6	54	1	V76140	Staphylococcus aur
C 19	13.6	48.6	69	1	V76230	Staphylococcus aur
C 20	13.4	47.9	97	1	Q06886	phcLSXVI mutant po
C 21	13.4	47.9	97	1	Q48454	phcLSXVI mutant po
C 22	13.2	47.1	23	1	T00694	Human trkC recepto
C 23	13.2	47.1	26	1	V53561	Nucleotide sequenc
C 24	13.2	47.1	29	1	Q25777	PCR primer CP19 to
C 25	13.2	47.1	29	1	Q29310	PCR primer CP19 to
C 26	13.2	47.1	29	1	Q35714	Canarypox ampliflc
C 27	13.2	47.1	29	1	Q67804	Canarypox primer,
C 28	13.2	47.1	29	1	O81594	Canarypox ppx4 PCR
C 29	13.2	47.1	29	1	T00930	Canarypox ppx4 PCR
C 30	13.2	47.1	29	1	T69968	Canarypox virus C3
C 31	13.2	47.1	29	1	T47576	Primer CP19 for 27
C 32	13.2	47.1	31	1	X38735	Human genomic DNA
C 33	13.2	47.1	35	1	V02251	Murine IgG PCRn ge
C 34	13.2	47.1	56	1	V66089	Synthetic fragment

C 35	13.2	47.1	59	1	V77728	Staphylococcus aur
C 36	13.2	47.1	72	1	T18145	Oligo used in synt
C 37	13.2	47.1	80	1	Q46877	Ovine GHG primer O
C 38	13.2	47.1	94	1	T33702	Human gene signatu
C 39	13	46.4	21	1	N82281	Fragment encoding
C 40	13	46.4	24	1	Q78344	Mumps haemagglutin
C 41	13	46.4	30	1	Q03277	Phe-355 mutation o
C 42	13	46.4	34	1	T70361	Rev/tax mRNA of bo
C 43	13	46.4	48	1	Q87345	Oligonucleotide pr
C 44	13	46.4	48	1	T13534	Test analyte from
C 45	13	46.4	55	1	X15963	PCR primer AL-7 us

## ALIGNMENTS

RESULT 1	
X22061/c	
ID X22061 standard; DNA; 24 BP.	
AC X22061:	
DT 19-MAY-1999 (first entry)	
DE PCR primer for C. maltosa cytochrome POX4 gene.	
KW PCR primer: POX4 gene; Pichia pastoris; di-carboxylic acid production;	
KW 6-22C mono-carboxylic acid production; alkane hydroxylation; ss.	
OS Synthetic.	
OS Candida maltosa.	
PN W09904014-A2.	
PD 28-JAN-1999.	
PR 20-JUL-1998; U14935.	
PR 21-JUL-1997; US-053215.	
PA (DUPLO ) DU POINT DE MEMOIRS & CO E. I.	
PI Fallon RD. Payne MS, Picatagyo SK, Wu S;	
PI WPI: 99-132258/11.	
DR Production of mono- and di-carboxylic alkanic acids in new	
PT engineered yeast - containing inserted copies of genes for	
PI cytochrome P450 monooxygenase and/or reductase, providing increased	
PI production and conversion efficiency, useful e.g. as intermediates	
PI for polymers and surfactants	
PS Example 6; Page 33; 57pp; English.	
CC This sequence represents a PCR primer for the POX4 gene	
CC of Candida maltosa. The invention relates to a method for the	
CC production of 6-22C mono- and di-carboxylic acids (I), which comprises	
CC treating, under aerobic conditions: (i) Pichia pastoris having	
CC genetically engineered alkane hydroxylation activity; or (ii) Candida	
CC maltosa genetically engineered to have enhanced alkane hydroxylation	
CC activity and/or a blocked beta-oxidation pathway, with a 6-22C linear	
CC hydrocarbon. The method is specifically used to make dodecanedioic acid	
CC (Ia) from dodecane, but generally any 6-22C alkane, or derived	
CC monocarboxylic acids or esters, can be converted. (I) that are	
CC monocarboxylic acids are intermediates for surfactants and those that are	
CC dicarboxylic acids are intermediates for polymers and anticorrosion	
CC agents. Transformation with the specified genes introduces (I)-producing	
CC ability to P. pastoris and enhances that in C. maltosa, resulting in	
CC conversion efficiency and productivity suitable for a commercial process	
CC (that is less environmentally damaging than known chemical syntheses),	
CC Disruption of the beta-oxidation pathway increases metabolite flow to the	
CC omega-oxidation pathway, and thus yield of, and selectively for, (I).	
SO Sequence 24 BP; 6 A; 4 C; 4 G; 10 T;	
Query Match 56.4%; Score 15.8; DB 1; Length 24;	
Best local Similarity 89.5%; Pred. No. 1.2e+02;	
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
OY 9 tgggtccaaatgtgtca 27	
DB 23 tgcgtacaaatgtgtca 5	
RESULT 2	
V43049/c	
ID V43049 standard; DNA; 29 BP.	
AC V43049:	

DT 23-OCT-1998 (first entry)  
DE 5' PCR primer Deg-A used to amplify nitric oxide synthase DNA.  
KW Nitric oxide synthase; treatment; epidermal; dermal condition;  
KM unmeted pigmentation; PCR primer; ss.  
OS Synthetic.  
OS Mus sp.  
OS Rattus sp.  
OS Homo sapiens.  
PN WO9833379-A1.  
PD 06-AUG-1998.  
PF 03-FEB-1998: U01891.  
PR 04-FEB-1997: US-037098.  
PA (GENO ) GEN HOSPITAL CORP.  
PI Lerner EA, Lerner LH, Qureshi AA;  
DR WPI: 98-437056/37.  
PT Treatment of unwanted epidermal or dermal conditions - comprising  
PT administration of treatment which modulates level of nitric oxide in  
PT skin  
PS Disclosure: Page 17, 35pp; English.  
CC PCR primers V43049-50 were used to amplify DNA encoding nitric oxide  
CC synthase (inducible isoform, the isoform found in nerve cells, and  
CC isoform found in epithelial cells). The specification describes  
CC a method for treating a subject for an unwanted epidermal or dermal  
CC condition, which comprises administration of a compound which modulates  
CC the level of nitric oxide in the skin. Where the condition is  
CC characterised by unwanted cells or unwanted pigmentation, the method  
CC includes increasing the level of nitric oxide in the skin. Where the  
CC condition is characterised by the lack of or by an insufficient number  
CC of dermal or epidermal cells or a lack of pigmentation, the method  
CC includes reducing the level of nitric oxide in the skin.  
SQ Sequence 29 BP; 12 A; 7 C; 5 G; 2 T;

Query Match 54.3%; Score 15.2; DB 1; Length 29;  
Best Local Similarity 77.3%; Pred. No. 2.4e+02;  
Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 ccctatgttggtgccaaattg 22  
||||| ||||| ||| |||  
DB 27 CCTTGTGTGGTGCGCRACATTG 6

RESULT 3  
ID Q71835 standard; DNA: 34 BP.  
AC Q71835;  
DT 30-MAR-1995 (first entry)  
DE Primer FIB2.  
KW Fibronectin binding protein; Fbp; fibronectin binding domain;  
KW monoclonal antibody; MAb; adhesion; Gram-positive bacterium;  
KW bacteremia; Escherichia coli; polymerase chain reaction; PCR;  
OS Amplification; primer; Staphylococcus aureus; ss.  
OS Synthetic.  
PN WO9418327-A.  
PD 18-AUG-1994.  
PF 04-FEB-1994: G00215.  
PR 05-FEB-1993: GB-002289.  
PR 20-OCT-1993: GB-021592.  
PA (SMIR ) SMITHKLINE BEECHAM PLC.  
PI Burnham MKR, Chopra I, Critchley IA, Knowles DJC;  
DR WPI: 94-279748/34.  
PT Fibronectin binding protein and monoclonal antibodies specific  
PT for Fbp - useful to prevent adherence of Gram-positive bacteria  
PT to indwelling devices or wounds  
PS Disclosure: Page 29; 40pp; English.  
CC Polypeptides corresponding to residues G709-T886 plus PPTVPT,  
CC G709-P838(P838) and G709-P838 (R58805-07, respectively) of  
CC S. aureus J2385 (NCIMB 40532) Fbp type A fibronectin binding  
CC domain D1-D4 region (R58808) were expressed in E. coli BL21(DE4)  
CC and used to raise MAbS specific for Fbp. DNA encoding  
CC D1-D4(G709-T886) was obtained by PCR amplification of S. aureus  
CC chromosomal DNA using primers FIB1 (Q71834) and FIB2 (Q71835).  
SQ Sequence 34 BP; 5 A; 6 C; 14 G; 9 T;

Query Match 53.6%; Score 15; DB 1; Length 34;  
Best Local Similarity 78.3%; Pred. No. 3e+02;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ccctatgttggtgccaaattg 23  
||| ||||| ||| |||||  
DB 7 CCTTACCTGTGGTGACAGATTGG 29

RESULT 4  
ID T18304 standard; DNA: 34 BP.  
AC T18304;  
DT 13-NOV-1996 (first entry)  
DE Primer for amplifying fibronectin binding domains.  
KW Fibronectin; oral hygiene; gingivitis; gum disease; adhesion;  
KW periodontitis; oral pathogen; caries; calculus; candidiasis;  
KW oral surgery; tissue regeneration; irrigation; toothpaste;  
KW dentifrice; mouthwash; lozenge; ss.  
OS Synthetic.  
PN WO9604003-A1.  
PD 15-FEB-1996.  
PF 18-JUL-1995: E02825.  
PR 05-AUG-1994: GB-015902.  
PA (SMIR ) SMITHKLINE BEECHAM PLC.  
PI Barnett P, Critchley IA, Dodd I;  
DR WPI: 96-129122/13.  
PT Prevention of adherence of oral pathogens in the oral cavity,  
PT partic. tooth surfaces - by application of a fibronectin binding  
PT protein or polypeptide or a monoclonal antibody or fragment against  
PT it  
PS Example 1: Page 18: 41pp; English.  
CC Adherence of oral pathogens, particularly to tooth surfaces, can be  
CC prevented by application of a fibronectin binding protein or  
CC polypeptide. The fibronectin binding protein or polypeptide is  
CC useful in the manufacture of oral hygiene compositions, eg.  
CC toothpaste, liquid dentifrice, mouthwash or lozenge. They are  
CC useful to prevent the plaque-related development of carious lesions,  
CC gingivitis, calculus or periodontal disease and to combat oral cavity  
CC infections, e.g. candidiasis. They may also be useful in oral  
CC surgery e.g. in guided tissue regeneration procedures to prevent  
CC subsequent bacterial infection, and for irrigation of periodontal  
CC pockets. Two primers (T18303, T18304) were used to amplify DNA  
CC encoding the fibronectin binding domains of fibronectin binding  
CC protein from the chromosomal DNA of Staphylococcus aureus J2385.  
SQ Sequence 34 BP; 5 A; 6 C; 14 G; 9 T;

Query Match 53.6%; Score 15; DB 1; Length 34;  
Best Local Similarity 78.3%; Pred. No. 3e+02;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ccctatgttggtgccaaattg 23  
||| ||||| ||| |||||  
DB 7 CCTTACCTGTGGTGACAGATTGG 29

RESULT 5  
ID X27477 standard; DNA: 54 BP.  
AC X27477;  
DT 08-JUN-1999 (first entry)  
DE Oligo 1131 for hCG alpha-subunit analogues.  
KW Analogue; heterodimeric; glycoprotein hormone; hCG; hLH; hFSH; hTSH;  
KW human chorionic gonadotropin; human luteinising hormone; disulphide bond;  
KW human follicle stimulating hormone; human thyroid stimulating hormone;  
OS Stability; primer; amplification; PCR; mutation; ss.  
OS Synthetic.  
OS Homo sapiens.  
PN WO9858957-A2.  
PD 30-DEC-1998.

```
PF 25-JUN-1998; U13070.
PR (ISTF ) ARS APPLIED RES SYSTEMS HOLDING NV.
PA (MCIN-) MCINNIS P G.
PI MOYLE WR:
DR WPI: 99-081219/07.
PT New stabilised glycoprotein hormones - particularly hCG, hLH, hFSH
PT or hTSH, have an intersubunit disulphide crosslink between the
PT alpha and beta subunits to improve stability
PS Example 1; Page 60; 139pp; English.
CC The invention relates to the production of analogues of a heterodimeric
CC subunit glycoprotein hormone (GPH) e.g. human chorionic gonadotropin
CC (hCG), human luteinising hormone (hLH), human follicle stimulating
CC hormone (hFSH), human thyroid stimulating hormone (hTSH), and functional
CC mutants, which are modified to contain an intersubunit disulphide bond,
CC between an alpha-subunit cysteine and a beta-subunit cysteine, for
CC improved stability, the analogue retaining at least a portion of the
CC bioactivity for the corresponding native GPH receptor. Primers
CC X27450-X27484 were used to introduce the mutations in the hCG sequence.
CC The improved analogues are designed specifically to reduce perturbation
CC of the 3-dimensional structure of the hormone, thereby creating greater
CC likelihood that the dimer will be formed in vivo and the formed dimer
CC will have affinity for the native receptors and have agonistic activity
CC on them. The changes stabilise the GPHs and prolong the biological
CC activities of the hormones. The analogues can have uses as for the
CC native GPHs.
SQ Sequence 54 BP; 18 A; 11 C; 12 G; 13 T;

Query Match 51.4%; Score 14.4; DB 1; Length 54;
Best Local Similarity 75.0%; Pred. No. 5.9e+02;
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 3 ctatgttggtgccaattgctca 26
    |||||
DB 20 CTATGTGTGTCCTCAAAAGACGTCA 43

RESULT 6
X27467
ID X27467 standard; DNA; 60 BP.
AC X27467;
DE 08-JUN-1999 (first entry)
KW Analogue; heterodimeric; glycoprotein hormone; hCG; hLH; hFSH; hTSH;
KW human chorionic gonadotropin; human luteinising hormone; disulphide bond;
KW human follicle stimulating hormone; human thyroid stimulating hormone;
KW stability; primer; amplification; PCR; mutation; ss.
OS Synthetic.
OS Homo sapiens.
PN WO9858957-A2.
PD 30-DEC-1998.
PF 25-JUN-1998; U13070.
PR 25-JUN-1997; US-050784.
PA (ISTF ) ARS APPLIED RES SYSTEMS HOLDING NV.
PI (MCIN-) MCINNIS P G.
PI MOYLE WR:
DR WPI: 99-081219/07.
PT New stabilised glycoprotein hormones - particularly hCG, hLH, hFSH
PT or hTSH, have an intersubunit disulphide crosslink between the
PT alpha and beta subunits to improve stability
PS Example 1; Page 60; 139pp; English.
CC The invention relates to the production of analogues of a heterodimeric
CC subunit glycoprotein hormone (GPH) e.g. human chorionic gonadotropin
CC (hCG), human luteinising hormone (hLH), human follicle stimulating
CC hormone (hFSH), human thyroid stimulating hormone (hTSH), and functional
CC mutants, which are modified to contain an intersubunit disulphide bond,
CC between an alpha-subunit cysteine and a beta-subunit cysteine, for
CC improved stability, the analogue retaining at least a portion of the
CC bioactivity for the corresponding native GPH receptor. Primers
CC X27450-X27484 were used to introduce the mutations in the hCG sequence.
CC The improved analogues are designed specifically to reduce perturbation
CC of the 3-dimensional structure of the hormone, thereby creating greater
CC likelihood that the dimer will be formed in vivo and the formed dimer
CC will have affinity for the native receptors and have agonistic activity
CC on them. The changes stabilise the GPHs and prolong the biological
CC activities of the hormones. The analogues can have uses as for the
CC native GPHs.
SQ Sequence 54 BP; 18 A; 11 C; 12 G; 13 T;
```

```
CC likelihood that the dimer will be formed in vivo and the formed dimer
CC will have affinity for the native receptors and have agonistic activity
CC on them. The changes stabilise the GPHs and prolong the biological
CC activities of the hormones. The analogues can have uses as for the
CC native GPHs.
SQ Sequence 60 BP; 18 A; 13 C; 14 G; 15 T;

Query Match 51.4%; Score 14.4; DB 1; Length 60;
Best Local Similarity 75.0%; Pred. No. 6e+02;
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 3 ctatgttggtgccaattgctca 26
    |||||
DB 20 CTATGTGTGTCCTCAAAAGACGTCA 43

RESULT 7
X27464
ID X27464 standard; DNA; 84 BP.
AC X27464;
DE 08-JUN-1999 (first entry)
DE Oligo 839 for hCG alpha-subunit analogues.
KW Analogue; heterodimeric; glycoprotein hormone; hCG; hLH; hFSH; hTSH;
KW human chorionic gonadotropin; human luteinising hormone; disulphide bond;
KW human follicle stimulating hormone; human thyroid stimulating hormone;
KW stability; primer; amplification; PCR; mutation; ss.
OS Synthetic.
OS Homo sapiens.
PN WO9858957-A2.
PD 30-DEC-1998.
PF 25-JUN-1998; U13070.
PR 25-JUN-1997; US-050784.
PA (ISTF ) ARS APPLIED RES SYSTEMS HOLDING NV.
PI (MCIN-) MCINNIS P G.
PI MOYLE WR:
DR WPI: 99-081219/07.
PT New stabilised glycoprotein hormones - particularly hCG, hLH, hFSH
PT or hTSH, have an intersubunit disulphide crosslink between the
PT alpha and beta subunits to improve stability
PS Example 1; Page 60; 139pp; English.
CC The invention relates to the production of analogues of a heterodimeric
CC subunit glycoprotein hormone (GPH) e.g. human chorionic gonadotropin
CC (hCG), human luteinising hormone (hLH), human follicle stimulating
CC hormone (hFSH), human thyroid stimulating hormone (hTSH), and functional
CC mutants, which are modified to contain an intersubunit disulphide bond,
CC between an alpha-subunit cysteine and a beta-subunit cysteine, for
CC improved stability, the analogue retaining at least a portion of the
CC bioactivity for the corresponding native GPH receptor. Primers
CC X27450-X27484 were used to introduce the mutations in the hCG sequence.
CC The improved analogues are designed specifically to reduce perturbation
CC of the 3-dimensional structure of the hormone, thereby creating greater
CC likelihood that the dimer will be formed in vivo and the formed dimer
CC will have affinity for the native receptors and have agonistic activity
CC on them. The changes stabilise the GPHs and prolong the biological
CC activities of the hormones. The analogues can have uses as for the
CC native GPHs.
SQ Sequence 84 BP; 25 A; 20 C; 16 G; 23 T;

Query Match 51.4%; Score 14.4; DB 1; Length 84;
Best Local Similarity 75.0%; Pred. No. 6.3e+02;
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 3 ctatgttggtgccaattgctca 26
    |||||
DB 44 CTATGTGTGTCCTCAAAAGACGTCA 67

RESULT 8
O97403/c
ID O97403 standard; DNA; 72 BP.
AC O97403;
```

DT	01-APR-1996 (first entry)	
DE	Mutated complementarity determining region L1.	
KW	Polymerase chain reaction; PCR; primer; amplification; recombination;	
KW	chimeric sequence; interleukin-1; antibody; tpa; growth hormone;	
KW	repetitive DNA; ribozyme; aptamer; gene therapy; A10B wild-type antibody;	
CS	complementarity determining region; ss.	
OS	Synthetic.	
EH	Key	Location/Qualifiers
FT	misc_difference 22. 51	
FT	/*tag= A	/note= "mutagenised region of wild type CDR H2 (see
FT	996186 for wild type)"	
PN	W09522625-A1.	
PD	24-AUG-1995.	
PF	17-FEB-1995; U02126.	
PR	17-FEB-1994; US-198431.	
PA	(AFFY-) AFFYMAX TECHNOLOGIES NV.	
PI	Cramer1 A, Stemmer WPC;	
DR	WPI: 95-302727/39.	
PT	DNA mutagenesis via random fragmentation and reassembly - useful in	
PT	the prodn. of mutant proteins having enhanced biological activity	
PS	Example 7; Page 82; 120pp; English.	
CC	The sequences shown in 097400-097405 are mutated versions of the	
CC	complementarity determining regions (CDRs) shown in 096183-096188. These	
CC	mutated CDRs were added to reassembly products of the A10B wild-type	
CC	antibody gene. The reassembly products were created by allowing the	
CC	50-200 bp fragments, produced upon DNaseI digestion, to reassemble. Once	
CC	the CDRs were added shuffling was allowed to occur. This method can also	
CC	be used to shuffle whole gene sequences. The method can also be used to	
CC	produce recombinant proteins, and chimeric DNA sequences. The methods	
CC	can be used to produce proteins with a desired phenotype having an	
CC	advantageous predetermined selectable property. Proteins such as	
CC	interleukin-1, antibodies, tpa and growth hormones can be generated that	
CC	have altered specificity or activity. Mutant sequences such as promoter	
CC	regions, introns, exons and enhancer sequences can also be generated,	
CC	which can be used to generate genes having increased expression rates.	
CC	The method can be used in the study of repetitive DNA regions, and for	
CC	the mutation of ribozymes or aptamers. Shifted viral proteins which may	
CC	comprise epitopes that will arise by viral evolution can also be	
CC	produced. This method of shuffling sequences may be used for generating	
CC	gene therapy vectors and replication-defective gene therapy constructs.	
CC	The advantage with this method is that it is not necessary to know the	
CC	actual DNA or RNA sequence of the specific nucleic acid fragment.	
SO	Sequence 72 BP; 21 A; 17 C; 18 G; 16 T;	
Query Match	50.7%; Score 14.2; DB 1; Length 72;	
Best Local Similarity	70.4%; Pred. No. 7.6e+02;	
Matches 19; Conservative 0; Mismatches 8; Indels 0; Gaps 0;		
Oy	2 cctatgttcgtccaaatggtcaca 28	
Db	65 cttctgttggtacacatattatctaacg 39	
RESULT 9		
ID	T73061/c	
AC	T73061 standard; DNA: 72 BP.	
AC	T73061.	
DT	13-FEB-1998 (first entry)	
DE	Mutant CDR oligonucleotide L1 for the murine antibody A10B scfv gene.	
KW	TEM-1 beta-lactamase gene; gene reassembly reaction; DNA library;	
KW	gene shuffling; random fragmentation; mutagenesis; recombination;	
KW	GFP protein; arsenate detoxification bacteria; cetotaxime; antibody A10B;	
KW	cadmium detoxification bacteria; drug resistance gene; ss.	
OS	Synthetic.	
PN	W09720078-A1.	
PD	05-JUN-1997.	
PF	02-DEC-1996; U19256.	
PR	25-MAR-1996; US-621859.	
PR	30-NOV-1995; US-564955.	
PA	(AFFY-) AFFYMAX TECHNOLOGIES NV.	
PI	Cramer1 A, Stemmer WPC;	

DR MPI: 97-310628/28.  
 PR Generating polynucleotide(s) with desired characteristics by  
 PR iterative selection and recombination - used for the directed  
 PR molecular evolution in vitro or in vivo of proteins, especially  
 PR green fluorescent protein  
 PS Example 7: Page 127; 209pp; English.  
 CC Sequences T73058-62 represent mutant complementarity determining regions  
 CC (CDRs) of the A10B scfv murine antibody gene. The present sequence  
 CC represents a CDR in the light chain. The original A10B antibody  
 CC reproducibility had only a low avidity, as it bound weakly to antigen. A  
 CC novel method was used to improve the antibody by DNA shuffling of a  
 CC library of all 6 mutant CDRs (based on CDRs T73046-51). The method  
 CC involves recombining at least 2 forms of a polynucleotide to  
 CC produce a library of recombinant forms of the sequence. The library is  
 CC screened for a recombinant sequence, which is then recombined with  
 CC another form of the polynucleotide, the same or different from the first  
 CC and second forms, to produce a further library of recombinant  
 CC polynucleotides. The subsequent libraries are screened, and recombination  
 CC performed until the recombinant polynucleotide has acquired the desired  
 CC property. The methods are used to reassemble DNA after random  
 CC fragmentation, for mutagenesis of nucleic acid sequences by in vitro or  
 CC in vivo recombination. The repeated cycles of mutagenesis, shuffling and  
 CC selection allow for the directed molecular evolution in vitro or in vivo  
 CC of proteins. In particular the methods are used for the production of  
 CC mutant GFP protein which has been modified to show enhanced fluorescence  
 CC when used as a reporter of gene expression and regulation. Also  
 CC exemplified is the use of the methods to generate improved arsenate  
 CC or cadmium detoxification bacteria, drug resistance genes and variant  
 CC polymerases.  
 SQ Sequence 72 BP; 21 A; 17 C; 18 G; 16 T;  
 Query Match 50.7%; Score 14.2; DB 1; Length 72;  
 Best Local Similarity 70.4%; Pred. No. 7, 66+02;  
 Matches 19; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 2 cctatgttgatgcccacaattggtcaag 28  
 | | | | | | | | | | | | | | | | | | | |  
 DB 65 cttctgttgatgacaaatgattgtatgtaacg 39  
 RESULT 10  
 VS4920/c  
 ID VS4920 standard; DNA; 72 BP.  
 AC VS4920;  
 DT 19-NOV-1998 (first entry)  
 DE Mutant light chain CDR L1 sequence.  
 KW A10B antibody gene; recombination; gene shuffling; ss.  
 OS Synthetic.  
 PN US5811238-A.  
 PD 22-SEP-1998.  
 PF 30-NOV-1995; 564955.  
 PR 04-MAR-1996; US-537874.  
 PR 17-FEB-1994; US-198431.  
 PR 30-NOV-1995; US-564955.  
 PA (AF97-) AFFYMAX TECHNOLOGIES NV.  
 PI Cramerl A. Stemmer WPC;  
 DR MPI: 98-530860/45.  
 PR Production of polynucleotides with desired properties - by iterative  
 PR selection and recombination  
 PS Example 7: Columns 55-56; 74pp; English.  
 CC The present sequence represents a mutant complementarity determining  
 CC region (CDR) L1 (see VS4926 for original sequence) from the light chain  
 CC of the A10B antibody gene. The antibody was improved by DNA shuffling of  
 CC a library of 6 mutant CDRs, using the method of the invention. The  
 CC specification describes a method for evolving a polynucleotide for  
 CC acquisition of a desired property. The method comprises providing a  
 CC population of variants of the polynucleotide, at least one of which is  
 CC in cell-free form, shuffling the variants of the polynucleotide to form  
 CC recombinant polynucleotides, selecting or screening for recombinant  
 CC polynucleotides that have evolved toward the desired property and  
 CC repeating the steps with the selected recombinant polynucleotides until  
 CC a recombinant polynucleotide has acquired the desired property. The

CC recombinant polynucleotide produced by method, or its expression product,  
CC can be formulated as a pharmaceutical.  
CC Sequence 72 BP; 21 A; 17 C; 18 G; 16 T;  
SQ

Query Match	50.7%	Score 14, 2;	DB 1;	Length 72;
Best Local Similarly	70.4%	Pred. No. 7.66+02;		
Matches 19; Conservative	0;	Mismatches 8;	Indels 0;	Gaps 0;

QY	2	cctatgttgggtgcacaaatgtgtcaag	28
Db	1		11
	2		12
	3		13
	4		14
	5		15
	6		16
	7		17
	8		18
	9		19
	10		20
	11		21
	12		22
	13		23
	14		24
	15		25
	16		26
	17		27
	18		28
	19		29
	20		30
	21		31
	22		32
	23		33
	24		34
	25		35
	26		36
	27		37
	28		38
	29		39
	30		40
	31		41
	32		42
	33		43
	34		44
	35		45
	36		46
	37		47
	38		48
	39		49
	40		50
	41		51
	42		52
	43		53
	44		54
	45		55
	46		56
	47		57
	48		58
	49		59
	50		60
	51		61
	52		62
	53		63
	54		64
	55		65
	56		66
	57		67
	58		68
	59		69
	60		70
	61		71
	62		72
	63		73
	64		74
	65		75
	66		76
	67		77
	68		78
	69		79
	70		80
	71		81
	72		82
	73		83
	74		84
	75		85
	76		86
	77		87
	78		88
	79		89
	80		90
	81		91
	82		92
	83		93
	84		94
	85		95
	86		96
	87		97
	88		98
	89		99
	90		100
	91		101
	92		102
	93		103
	94		104
	95		105
	96		106
	97		107
	98		108
	99		109
	100		110
	101		111
	102		112
	103		113
	104		114
	105		115
	106		116
	107		117
	108		118
	109		119
	110		120
	111		121

RESULT	11
V06529	
ID	V06529 standard; DNA; 27 BP.

DT 11-MAY-1998 (first entry)  
DE Equine chorionic gonadotropin alpha-subunit cDNA mutagenic primer 1.  
KW Equine; Chorionic gonadotropin; ecg; alpha-subunit; hormone; variant;  
KW lutealising hormone; treatment; ovarian disease; ovulation; mutagenic;  
PCR primer; ss.

OS Family Equidae.  
PN J10036285-A.  
PD 10-FEB-1998.  
PR 23-JUL-1996; 1993232.  
PF 23-JUL-1996; JP-193232.  
PA (ELED) DENKI KAGAKU KOGYO KK

PT Mutant equine chorionic gonadotropin hormone - has substitution in  
PT alpha chain preventing binding of sugar to residue 56, resulting in  
PT loss of luteinising hormone activity  
NS Example 1; Pages 5-6; 15pp; Japanese.

CC This primer is used for the PCR amplification of a cDNA encoding an  
CC alpha-subunit of equine chorionic gonadotropin (ecg). This is used to  
CC create a variant of the alpha-subunit by PCR mutagenesis. The variant has  
CC an amino acid substitution to inhibit binding of a sugar chain to 56th  
CC amino acid resulting in loss of luteinizing hormone (LH) activity. A  
CC variant gonadotropin hormone composed of the variant alpha-subunit and  
CC beta-subunit of ecg is a safe and widely applicable ovulation inducer and  
CC may be used as an agent for the treatment of ovarian diseases. This  
CC gonadotropin hormone is substantially free from LH activity.  
Sequence 27 BP: 8 A; 8 C; 5 G; 6 T;

Query Match	50.08;	Score 14;	DB 1;	Length 27;
Best Local Similarity	77.38;	Pred. No. 8e+02;		
Matches 17; Conservative	0;	Mismatches 5;	Indels 0;	Gaps 0;

```

QY      5 atgttgtgccaatgtgtca 26
          |||||  |||||  |||
Db      1 ATGTTGGTCCCAAGCAGATCA 22

```

## RESULT 12

ID Q52717 standard; cDNA; 35 BP.

DT	24-JUN-1994 (first entry)
DE	Sequence of PCR primer C185-A

KW	PCR primer; multi-drug resistant gene; mdrl; plasmid G1MD1; ss.
OS	Synthetic.
PN	W09324613-A.

PD	09-DEC-1993.
PF	14-MAY-1993; U04707.
PR	22-MAY-1992; US-887712.

PA (GENE-) GENETIC THERAPY INC.  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PI Mcdonagh KT, Nienhuis A, Tolstoshev P;

PT DNA or RNA sequence for human multi-drug resistant gene MDR1 -

PT encodes P-glycoprotein with altered splice site, used in cancer therapy  
 PT Example; page 18; 64pp; English.

CC G6IMD1 contains the mdri gene. The cryptic splice donor site and the  
CC G1Y to Val point mutation in codon 185 are contd. Within a NotI to  
CC Apal restriction fragment of g6im1. Following subcloning of the  
CC NotI to Apal fragment into pUC007 (to form pMD1NA), the cryptic  
CC splice donor site and the codon 185 point mutation will be contd.  
CC within a unique BstXI to MscI fragment. A strategy of overlapping  
CC PCR was used to introduce point mutations into this BstXI to MscI  
CC fragment. Three sets of PCR primers are made. One set of  
CC complementary primers (SD-A and SD-B) (952715, 952716) are centred

CC around the cryptic splice donor site, and differ from the sequence  
CC of GIM1 at only one base, located in codon 139 of the *mrl* gene.  
CC wherein AGG is changed to AGA. This difference inactivates the  
CC splice donor site, although the AA is not changed. The second set  
CC of primers (C185-A and C185-B) (052717, 052718) are centred at codon

CC 185 mutation of the mdrl gene and will differ from the sequence of  
CC G14M1 only at the point mutation wherein codon 185 is changed from  
CC GGT to GGT. This difference corrects the point mutation in the cDNA  
CC sequence and encodes Gly in the final construction. The third set of  
CC primers flank the BstXI and MscI restriction sites. These primers  
CC 5'Bst and 3'Msc (Q52719, Q52720) overlap the BstXI and MscI sites,  
CC

CC respectively, and also incorporate another flanking restriction site  
CC that will allow the PCR fragment to be conveniently subcloned.  
CC Using the technique of recombinant PCR three separate PCR reactions  
CC will be initiated, using mDNA as the template DNA.  
SQ Sequence 35 BP; 13 A; 3 C; 8 G; 11 T;

Query Match	49.3%;	Score 13.8;	DB 1;	Length 35;
Best Local Similarity	88.2%;	Pred. No. 1e+03;		
Matches 15; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

Qy	6	tg	tg	tg	cc	aa	at	tg	22
Db	19	TAT	TGG	TGAC	CAAA	AT	TG		35

## RESULT

ID Q52718 standard; cDNA; 35 BP.

DT 24-JUN-1994 (first entry)

KW PCR primer;

PN W09324613-A.

PF 14-MAY-1993;

## PA (GENE-) GENETIC THERAPY

PI Mcdonagh KT, Nienhuis A, Tolstoshev P.

[illegible]

PT	therapy	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
PT	therapy	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

CC PG1MD1 contains the *mdr1* gene. T

CC ApaI restriction fragment of pG1MD1. Following subcloning of the

CC splice donor site and the codon 185 point mutation will be contd.

CC PCR was used to introduce point mutations into this BstXI to MscI

CC complementary primers (SD-A and SD-B) (Q52715, Q52716) are centred

CC of GIMD1 at only one base, located in codon 139 of the mdr1 gene,

CC splice donor site, although the AA is not changed. The second set





Tue Jun 6 08:34:50 2000

us-09-164-714-8.rng

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 13:53:05 ; Search time 4521.53 Seconds  
(without alignments)  
25.100 Million cell updates/sec

Title: US-09-164-714-8

Perfect score: 28

Sequence: 1 ccctatgtgtgcacaaattggtcaag 28

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 202611650 residues

Total number of hits satisfying chosen parameters: 119968

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST.\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result

Query

No.	Score	Match	Length	DB	ID	Description
1	15.6	55.7	67	34	AA474067	AA474067 ves4e02.x
2	15.2	54.3	46	43	AI209047	AI209047 qg28e02.x
3	15	53.6	73	51	AI739189	AI739189 w127c11.x
4	15	53.6	82	51	AI745836	AI745836 tr04g11.x
5	15	53.6	89	23	H44598	H44598 ypl7c09.t1
6	15	53.6	93	44	AI282700	AI282700 qt83h07.x
7	15	53.6	97	37	AA714476	AA714476 nw06f12.s
8	14.8	52.9	97	35	AA566933	AA566933 987 l0bio
9	14.8	52.9	87	60	AI795116	AI795116 sb76h05.y
10	14.6	52.1	55	46	AI444423	AI444423 fb26b06.x
11	14.6	52.1	59	62	AI188047	AI188047 at78d09.x
12	14.6	52.1	92	23	H40392	H40392 ype0e01.t1
13	14.6	52.1	99	70	AM145122	AM145122 ga29e08.y
14	14.4	51.4	79	29	AA165762	AA165762 ms60f11.r
15	14.4	51.4	91	81	AM394554	AM394554 sb32e03.y
16	14.4	51.4	96	44	AI314653	AI314653 u37a11.x
17	14.4	51.4	98	41	AU008426	AU008426 AU008426
18	14.2	50.7	98	28	AA107577	AA107577 mp05a11.r
19	14.2	50.7	97	29	AA180917	AA180917 zp44b01.r
20	14	50.0	46	40	AA95439	AA95439 os84f01.s
21	14	50.0	88	47	AI474685	AI474685 tm35f12.x
22	14	50.0	91	36	AA659101	AA659101 nu81e01.s
23	13.8	49.3	86	29	AA146242	AA146242 mq88f06.r
24	13.8	49.3	89	26	W76217	W76217 zd58f09.t1
25	13.8	49.3	92	46	AI423670	AI423670 tf85c04.x
26	13.8	49.3	99	41	AU007058	AU007058 AU007058
27	13.6	48.6	56	84	B03327	B03327 CSR1-176D2-
28	13.6	48.6	61	38	AA746401	AA746401 nw62c10.s
29	13.6	48.6	61	38	AI338554	AI338554 qg93c08.x
30	13.6	48.6	88	24	N25181	N25181 yy02b10.s1
31	13.6	48.6	91	34	AA472378	AA472378 vho6g06.r
32	13.6	48.6	95	40	AA910572	AA910572 ok66h10.s
33	13.6	48.6	97	44	AI285065	AI285065 qk56f09.x
34	13.6	48.6	98	29	AA187895	AA187895 zp74g12.r
35	13.6	48.6	98	50	F29909	F29909 HSPD20053.H
36	13.4	47.9	58	41	AU007255	AU007255 AU007255
37	13.4	47.9	64	46	AI442743	AI442743 sa85h01.y
38	13.4	47.9	67	30	AA232723	AA232723 ar75e10.r
39	13.4	47.9	67	37	AA669910	AA669910 ag42g07.s
40	13.4	47.9	68	37	AA667085	AA667085 vr87c12.s
41	13.4	47.9	79	33	AA446093	AA446093 zw58a10.s
42	13.4	47.9	98	46	AI427907	AI427907 ml38e10.x
43	13.4	47.9	100	29	AA152082	AA152082 z148g12.r
44	13.2	47.1	34	45	AI367088	AI367088 qg46d03.x
45	13.2	47.1	49	24	H97554	H97554 yw04h01.s1

## ALIGNMENTS

RESULT 1  
 LOCUS AA474067 67 bp mRNA EST 18-JUN-1997  
 DEFINITION ves4e02.r1 Beddington mouse embryonic region Mus musculus cDNA  
 ; clone IMAGE:821978 5' similar to TR:G476095 G476095 B4-2 PROTEIN.  
 ; mRNA sequence.

ACCESSION AA474067  
 VERSION AA474067.1 GI:22022294  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 67)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theisinger,B., Wylie,T., Lennon,G., Soares,B., Willson,R. and  
 Waterston,R.

TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)

## COMMENT

On May 9, 1995 this sequence version replaced gi:803068.  
 Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LBNL; contact the  
 IMAGE Consortium (info@lmbl.llnl.gov) for further information.  
 MGI:490258  
 Trace considered overall poor quality  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -28ml3 rev1 ET from Amersham  
 High quality sequence stop: 1.  
 Location/Qualifiers  
 1..67  
 /organism="Mus musculus"  
 /strain="C57BL6 x DBA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:821978"  
 /clone\_lib="Beddington mouse embryonic region"  
 /sex="pooled"  
 /tissue\_type="embryo"  
 /dev\_stage="7.5dpc"  
 /lab\_host="DH12S"  
 /note="Organ: whole embryo; Vector: PCMV-SPORT; Site:1:  
 SalI; Site:2: NotI; Cloned unidirectionally. Primer:  
 oligo dt. Gastrulating embryos were collected at 7.5dpc  
 from C57BL6 x DBA matings, excluding embryos that had  
 developed head folds and all extraembryonic tissues.  
 Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).  
 Referenced in Development 121, 2479-2489 (1995)"

## BASE COUNT

19 a 23 c 13 g 12 t

## Query Match

Best local Similarity 81.8%; Score 15.6; DB 34; Length 67;  
 Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 ctatgtgtgtccaaattgt 24  
 ||||| ||||| ||||| |||||  
 Db 4 CTATGCTGTGAGCCAAATTAGT 25

RESULT 2  
 LOCUS AI209047 46 bp mRNA EST 29-NOV-1998  
 DEFINITION qg28e02.x1 NCI\_CGAP\_K1d3 Homo sapiens cDNA clone IMAGE:1762394 3'  
 similar to SW:Y124\_HUMAN Q14137 HYPOPHYRETICAL PROTEIN KIA0124 ;  
 ; mRNA sequence.

ACCESSION AI209047  
 VERSION AI209047.1 GI:3770989  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 46)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index

JOURNAL Unpublished (1997)  
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2151684.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher Koskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.llnl.gov/dbdp/image/image.html](http://www.bio.llnl.gov/dbdp/image/image.html)

Trace considered overall poor quality  
Insert Length: 692 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence step: 1.

## FEATURES

source

1. .46

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1762394"

/clone\_lib="NCI-CGAP\_K1d3"

/lab\_host="DH10B"

/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I - oligo(dT) primer, strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 10 a 9 c 17 g 10 t

## ORIGIN

## Query Match

Best Local Similarity 54.3%; Score 15.2; DB 43; Length 46;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 9 tgggtgccaaatggtcgaag 28

||||| ||||| ||||| |||||

Db 18 TGGTGCMAAAGTGTGAAG 37

## RESULT 3

AI739189

LOCUS

DEFINITION

AI739189 73 bp mRNA EST 18-JUN-1999  
w127c11.x1 NCI-CGAP\_Col6 Homo sapiens cDNA clone IMAGE:2391476 3'  
similar to gb:K17206 40S RIBOSOMAL PROTEIN S4 (HUMAN);, mRNA  
sequence.

AI739189 GI:5101170

VERSION

KEYWORDS

SOURCE

ORGANISM

human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)  
On Jun 5, 1998 this sequence version replaced gi:3189015.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Ilian Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.llnl.gov/dbdp/image/image.html](http://www.bio.llnl.gov/dbdp/image/image.html)

## FEATURES

source

Seq primer: -40UP from Gibco.

Location/Qualifiers

1. .73

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2391476"  
/clone\_lib="NCI-CGAP\_Col6"  
/tissue\_type="colon tumor, RER"

/lab\_host="DH10B"

/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; plasmid DNA from the normalized library NCI-CGAP\_Col6 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneds 1057416-1061255, and 114584-1145351).  
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 17 a 18 c 13 g 25 t

## ORIGIN

## Query Match

Best Local Similarity 53.6%; Score 15; DB 51; Length 73;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 cccatgttggtgccaaatgg 23

||||| ||||| ||||| |||||

Db 36 CCGATGTGTGACACAGCTGG 58

## RESULT 4

AI745436

LOCUS

DEFINITION

AI745436 82 bp mRNA EST 21-JUN-1999  
tr04q11.x1 NCI-CGAP\_Ov23 Homo sapiens cDNA clone IMAGE:2217380 3'  
similar to gb:K17206 40S RIBOSOMAL PROTEIN S4 (HUMAN);, mRNA  
sequence.

AI745436 GI:5113724

VERSION

KEYWORDS

SOURCE

ORGANISM

human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)  
On May 18, 1998 this sequence version replaced gi:3137118.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskalkuk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Washington University Genome Sequencing Center  
clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.llnl.gov/dbdp/image/image.html](http://www.bio.llnl.gov/dbdp/image/image.html)

## FEATURES

source

Seq primer: -40UP from Gibco.

Location/Qualifiers

1. .82

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2217380"

/clone\_lib="NCI-CGAP\_Ov23"

/tissue\_type="tumor, 5 pooled (see description)"

/lab\_host="DH10B"

/note="Organ: ovary; Vector: PCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.35 kb. Tumor types include: mixed Mullerian tumor, papillary serous, clear cell, spindle cell. All are primary tumors, metastasis positive. Life Technologies catalog #: 11534-013"

BASE COUNT 17 a 15 c 10 g 40 t

## ORIGIN

Query Match 53.6%; Score 15; DB 51; Length 82;  
Best Local Similarity 78.3%; Pred. No. 66+03;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 cccatgtgtgtgccaattgg 23  
|||||  
Db 54 CCTATGTTGTAGCCACAGCTGG 76

## RESULT 5

LOCUS H44598 89 bp mRNA EST 31-JUL-1995  
DEFINITION ypl7c09.r1 Soares breast 3BDHbst Homo sapiens cDNA clone  
IMAGE:187696 5' similar to gb:W80563 PLACENTAL CALCIUM-BINDING  
PROTEIN (HUMAN);, mRNA sequence.

ACCESSION H44598.1 GI:920650  
VERSION H44598.1  
KEYWORDS EST.  
SOURCE human.

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 89)

## REFERENCE

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.

The WashU-Merck EST Project  
Unpublished (1995)  
Contact: Wilson RK

## JOURNAL

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu

Insert Size: 585  
High quality sequence starts: 1  
High quality sequence stops: 1

Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Trace considered overall poor quality  
Insert length: 585 Std Error: 0.00  
Seq primer: M13Rev

High quality sequence stop: 1.  
Location/Qualifiers

1. 89  
/organism="Homo sapiens"  
/db\_xref="GDB:3818592"  
/db\_xref="taxon:9606"  
/clone="IMAGE:187696"  
/clone\_lib="Soares breast 3BDHbst"  
/sex="Female"  
/dev\_stage="adult"  
/lab\_host="PH10B (ampicillin resistant)"  
/note="Organ: breast; Vector: pTZ19D (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - Oligo(dT) primer [5'  
TGTTCACATCTGAAGTGGACGGCCGCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of a modified pTZ19 vector (Pharmacia).  
Library went through one round of normalization to a Cot =  
20. Library constructed by Bento Soares and M.Fatima  
Bonaldo."

## FEATURES

source

BASE COUNT 22 a 19 c 15 g 23 t 10 others  
ORIGIN

Query Match 53.6%; Score 15; DB 23; Length 89;  
Best Local Similarity 72.0%; Pred. No. 6.1e+03;  
Matches 18; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 cccatgtgtgtgccaattggc 26  
|||||  
Db 38 CCTTGTGTGTCACACAGCTGA 14

## RESULT 6

LOCUS A1282700 93 bp mRNA EST 23-NOV-1998  
DEFINITION qc83h07.x1 NCI-CGAP\_Col4 Homo sapiens cDNA clone IMAGE:1961917 3'  
similar to gb:X17206 40S RIBOSOMAL PROTEIN S4 (HUMAN);, mRNA  
sequence.

ACCESSION A1282700  
VERSION A1282700.1 GI:3920933  
KEYWORDS EST.  
SOURCE human.

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 93)

## REFERENCE

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

## JOURNAL

Unpublished (1997)  
On Jan 17, 1998 this sequence version replaced gi:1900073.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.

cDNA library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
[www-bio.lnl.gov/dbp/image/image.html](http://www-bio.lnl.gov/dbp/image/image.html)

Seq primer: -40UP from Gibco.

## FEATURES

source

1. 93  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1961917"  
/clone\_lib="NCI-CGAP\_Col4"  
/tissue\_type="moderately-differentiated adenocarcinoma"  
/lab\_host="PH10B"  
/note="Organ: colon; Vector: PCMV-SPORE6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.7 kb. Life Technologies catalog #:  
11531-019"

BASE COUNT 20 a 22 c 21 g 30 t  
ORIGIN

Query Match 53.6%; Score 15; DB 44; Length 93;  
Best Local Similarity 78.3%; Pred. No. 6.1e+03;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 cccatgtgtgtgccaattgg 23  
|||||  
Db 37 CCTATGTTGTAGCCACAGCTGG 59

## RESULT 7

LOCUS AA714476 97 bp mRNA EST 22-JAN-1998  
DEFINITION nw06f12.s1 NCI-CGAP\_S51 Homo sapiens cDNA clone IMAGE:1238543 3'  
similar to gb:X17206 40S RIBOSOMAL PROTEIN S4 (HUMAN);, mRNA  
sequence.  
ACCESSION AA714476

```

VERSION AA714476.1 GI:2726750
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eukarya; Primates; Catarrhini; Hominiidae; Homo.
TITLE 1 (bases 1 to 97)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1407494.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmett-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html

Insert Length: 771 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.
Location/Qualifiers
1. 97
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1238543"
/clone_lib="NCI-CGAP-S61"
/tissue_type="synovial sarcoma"
/lab_host="SOLR (kanamycin resistant)"
/Note="Vector: Bluescript SK-; Site.1: EcoRI; Site.2:
XhoI; Cloned unidirectionally. Primer: Oligo dT. Synovial
sarcoma. 5' adaptor sequence: 5' GAATCGGCGACGAG 3'
adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' Average
insert size: 1.4 kb."

BASE COUNT 18 a 24 c 26 g 29 t
ORIGIN

Query Match 53.6%; Score 15; DB 37; Length 97;
Best Local Similarity 78.3%; Pred. No. 6.2e+03;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 cccatgttggtgccaaattgg 23
||||| 11111 1111 111
Db 25 CCCATGTGTGTAGCCACACGCTGG 47

RESULT 8
AA566933 59 bp mRNA EST 28-AUG-1998
LOCUS 987 lobiolly pine CA Pinus taeda CDNA clone 1CA12G, mRNA sequence.
AA566933
VERSION AA566933.1 GI:3366146
KEYWORDS EST.
SOURCE lobiolly pine.
ORGANISM Pinus taeda
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Coniferopsida; Coniferales;
Pinaceae; Pinus.
REFERENCE 1 (bases 1 to 59)
AUTHORS Allona, T., Quinn, M., Shoop, E., Swope, K., St. Cyr, S., Carlis, J.,
Riedl, J., Retzel, E., Campbell, M., Sederoff, R. and Whetten, R.W.
TITLE Analysis of xylem formation in pine by cDNA sequencing
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (16), 9693-9698 (1998)
MEDLINE 96356220
COMMENT On Sep 12, 1996 this sequence version replaced gi:1394808.
Contact: Ross Whetten

```

```

FOREST Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall,
Raleigh, NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhetten@uncsu.edu
Seq primer: T3.
Location/Qualifiers
1. 59
/organism="Pinus taeda"
/strain="Coastal plain lobiolly pine from North Carolina"
/db_xref="taxon:3352"
/clone="1CA12G"
/clone_lib="lobiolly pine CA"
/tissue_type="Xylem"
/lab_host="SOLR"
/Note="Vector: lambda-ZAP; Site.1: EcoRI; Site.2: XhoI;
The result of subtraction of C library with N library.
Immature xylem from the underside of inclined stems of
differentiating compression wood was subtracted with
immature xylem from the side of inclined stems of
differentiating wood. A mixture of four genotypes were
used. Oligo-dT primed cDNA was directionally cloned into
the EcoRI-XhoI lambda-ZAP vector arms"

BASE COUNT 15 a 12 c 16 g 14 t
ORIGIN

Query Match 52.9%; Score 14.8; DB 35; Length 59;
Best Local Similarity 84.2%; Pred. No. 7e+03;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 9 tgggtccaaattggtcaa 27
||||| 11111 11111
Db 14 TGGTGCCCATGATGATCA 32

RESULT 9
A1795116 87 bp mRNA EST 13-DEC-1999
LOCUS sb76h05.y1 Gm-c1010 Glycine max CDNA clone GENOME SYSTEMS CLONE ID:
Gm-c1010-946 5', mRNA sequence.
A1795116
VERSION A1795116.1 GI:5342832
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
Glycine.
REFERENCE 1 (bases 1 to 87)
AUTHORS Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V.,
Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J.,
Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M.,
Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N.,
Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,
McCann, R., Waterston, R. and Wilson, R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT On Jun 22, 1998 this sequence version replaced gi:3247195.
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or

```







/lab\_host="DH10B"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Construction of the cDNA library was carried out using Stratagene's 'unizap' - cDNA synthesis kit'. cDNA was constructed using an oligo dt primer/linker that contains a XhoI site within it. Following ds cDNA synthesis, EcoRI adapters were ligated to the blunt ends and sample was digested with XhoI. The result is cDNA with an EcoRI sticky end on one side and a XhoI sticky end on the other. This cDNA was ligated directionally in unizap arms. The vector is designed containing the pBluescript sequence as well as lambda DNA and cDNA is cloned within this pBluescript sequence. The vector was then packaged using gold gigapackaging extracts. Library was grown in XLBlue MRF' cells and amplified. The library was excised by mass excision using Stratagene's 'Mass excision kit' that uses exsist as a helper phage that releases the pBluescript sequence and circularises it as single stranded plasmids that are then packaged (by helper phage) and secreted out of the host cell as phagemids. SOLR cells were transformed with phagemids and the library was plated out on LB-amp plates to select for transformants. Approximately 1,000,000 colonies were grown and recovered. The double stranded plasmid library was recovered by using Qiaagen Midi prep kit. 2 micro grams of each library were used to transform DH10B cells by electroporation."

BASE COUNT  
 ORIGIN  
 27 a 23 c 24 g 25 t

Query Match 52.1%; Score 14.6; DB 70; Length 99;  
 Best Local Similarity 81.0%; Pred. No. 9.4e+03;  
 Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 OY 8 ttgtgtccaaattgtcgaag 28  
 |||| |||| |||| |||| ||||  
 DB 89 TTGGACCGCAATTGACCAAG 69

RESULT 14  
 AA165762/c  
 LOCUS  
 DEFINITION  
 AA165762 79 bp mRNA EST 12-FEB-1997  
 ms60f11.r1 Stratagene mouse embryonic carcinoma (#337317) Mus musculus cDNA IMAGE:615961 5' similar to TR:E93245 E93245 ETN  
 INSERT IN THE P4S APOPTOSIS GENE OF MRL-IPR/IPR. [1] ; mRNA  
 sequence.  
 ACCESSION  
 AA165762  
 VERSION  
 AA165762.1 GI:1743977  
 EST.  
 SOURCE  
 house mouse.  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 79)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.  
 TITLE  
 The Washu-HMT Mouse EST Project  
 JOURNAL  
 Unpublished (1996)  
 COMMENT  
 On Sep 12, 1996 this sequence version replaced gi:1325676.  
 Contact: Marra M/Mouse EST Project  
 Washu-HMT Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:376805  
 Seq primer: -28m13 rev1 ET from Amersham

FEATURES  
 source  
 High quality sequence stop: 75.  
 Location/Qualifiers  
 1..79  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_1b="IMAGE:615961"  
 /clone\_1lb="Stratagene mouse embryonic carcinoma (#337317)"  
 /tissue\_type="carcinoma"  
 /dev\_stage="embryonic"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. p19 cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGACAG 3' -3' adaptor sequence: 5' CTCGACTTTT TTTT TTTT TTTT 3' "

BASE COUNT  
 ORIGIN  
 20 a 19 c 26 g 14 t

Query Match 51.4%; Score 14.4; DB 29; Length 79;  
 Best Local Similarity 75.0%; Pred. No. 1.1e+04;  
 Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 OY 1 ccctatgttgcgaattgt 24  
 ||| | ||| |||| |||| |  
 DB 56 CCCAGTTCTGCGCCCAAAATTGTT 33

RESULT 15  
 AW394554  
 LOCUS  
 DEFINITION  
 AW394554 91 bp mRNA EST 07-FEB-2000  
 sh32e03.y1 Gm-c1017 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
 Gm-c1017-3413 5', mRNA sequence.  
 ACCESSION  
 AW394554  
 VERSION  
 AW394554.1 GI:6912966  
 EST.  
 KEYWORDS  
 soybean.  
 SOURCE  
 Glycine max  
 ORGANISM  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eudicots I; Fabales; Fabaceae; Papilionoideae; Glycine.  
 1 (bases 1 to 91)  
 Shoemaker,R., Kelm,P., Vodkin,L., Erpelting,J., Corryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurr,K., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.  
 Public Soybean EST Project  
 Unpublished (1999)  
 On Jul 9, 1999 this sequence version replaced gi:5433560.  
 Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com  
 Seq primer: -40RP from Gibco.  
 Location/Qualifiers  
 1..91  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1017-3413"  
 /clone\_1lb="Gm-c1017"  
 /tissue\_type="vegetable buds of field grown plants"

FEATURES  
 source

```
/lab_host="XL10-Gold"
/Note="Vector: Bluescript II XR. Site_1: EcoRI; Site_2:
XhoI; This cDNA library was constructed from mRNA isolated
from vegetable buds of field grown plants. The cDNA
library was prepared using the Stratagene Bluescript II
XR library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dT) sequence with a XhoI restriction site. EcoRI adapters
were ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
Bluescript vector. The ligated cDNA fragments were
transformed into XL10-Gold host cells. This library was
constructed by Dr. Randy Shoemaker and Dr. John
Erpelidg."
BASE COUNT      22 a      33 g      25 t
ORIGIN          11 c
```

```
Query Match      51.4%; Score 14.4; DB 81; Length 91:
Best Local Similarity 75.0%; Pred. No. 1.1e+04;
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
QY      1 ccctatgtgtgtgccaaattggt 24
      ||||| ||||| |||||
Db      44 CCCTTGTGTGTCATATAATGCTT 67
```

```
Search completed: June 4, 2000, 13:53:09
Job time: 20933 sec
```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:09:40 ; Search time 244.64 Seconds  
(without alignments)  
14.877 Million cell updates/sec

Title: US-09-164-714-8

Perfect score: 28  
Sequence: 1 cccatgtgtgtgcacaaatgtgtcaag 28

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 375880

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/prodata/2/lna/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/2/lna/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/2/lna/5C.COMB.seq:\*  
4: /cgn2\_6/prodata/2/lna/5D.COMB.seq:\*  
5: /cgn2\_6/prodata/2/lna/6.COMB.seq:\*  
6: /cgn2\_6/prodata/2/lna/PCUTUS.COMB.seq:\*  
7: /cgn2\_6/prodata/2/lna/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	57.1	39	1	US-08-361-920-32
2	16	57.1	39	1	US-08-479-939-32
3	16	57.1	39	2	US-08-483-432-32
4	15	53.6	34	4	US-08-459-135A-2
5	14.2	50.7	72	1	US-08-198-431-31
6	14.2	50.7	72	2	US-08-564-955-31
7	14.2	50.7	72	3	US-08-537-874-31
8	14	50.0	36	2	US-08-764-100-6
9	13.8	49.3	35	3	US-08-583-276-9
10	13.8	49.3	35	3	US-08-583-276-10
11	13.6	48.6	21	4	US-08-743-637B-210
12	13.6	48.6	26	2	US-08-538-875-26
13	13.4	47.9	18	5	US-09-289-376-21
14	13.4	47.9	44	2	US-08-592-406-5
15	13.4	47.9	97	1	US-07-635-561A-3
16	13.2	47.1	23	3	US-08-359-705B-11
17	13.2	47.1	23	3	US-08-286-846A-11
18	13.2	47.1	23	4	US-08-457-880A-11
19	13.2	47.1	23	5	US-08-444-622A-11
20	13.2	47.1	23	5	US-08-942-562-11
21	13.2	47.1	29	1	US-08-105-483-268
22	13.2	47.1	29	2	US-08-224-391-84
23	13.2	47.1	29	2	US-08-484-304-84
24	13.2	47.1	29	2	US-08-224-657-105
25	13.2	47.1	29	2	US-08-709-209-268
26	13.2	47.1	29	2	US-08-257-073-63
27	13.2	47.1	29	2	US-08-458-101-268

c 28	13.2	47.1	29	3	US-08-184-009-134	Sequence 134, App
c 29	13.2	47.1	29	3	US-08-566-398-45	Sequence 45, App
c 30	13.2	47.1	29	4	US-08-458-356-134	Sequence 134, App
c 31	13.2	47.1	29	4	US-08-658-665-88	Sequence 88, App
c 32	13.2	47.1	43	1	US-07-885-689A-24	Sequence 24, App
c 33	13.2	47.1	43	4	US-08-857-946-116	Sequence 116, App
c 34	13.2	47.1	43	5	US-08-970-740-116	Sequence 116, App
c 35	13	46.4	21	1	US-08-434-411-38	Sequence 38, App
c 36	13	46.4	21	2	US-08-434-402-38	Sequence 38, App
c 37	13	46.4	21	2	US-08-783-288-38	Sequence 38, App
c 38	13	46.4	21	4	US-08-890-640-38	Sequence 38, App
c 39	13	46.4	21	7	5194592-69	Sequence 38, App
c 40	13	46.4	30	6	PCT-US93-01598-27	Sequence 27, App
c 41	13	46.4	48	1	US-08-116-389-16	Sequence 16, App
c 42	13	46.4	48	2	US-08-708-431-16	Sequence 16, App
c 43	13	46.4	48	3	US-08-880-830-16	Sequence 16, App
c 44	13	46.4	48	6	PCT-US94-13895-16	Sequence 16, App
c 45	13	46.4	56	6	PCT-US91-05177-18	Sequence 18, App

## ALIGNMENTS

RESULT 1  
US-08-361-920-32/C  
Sequence 32, Application US/08361920  
Patent No. 5457046  
GENERAL INFORMATION:  
APPLICANT: Woelldike, Helle F.  
APPLICANT: Hagen, Frederick  
APPLICANT: Hjort, Carsten M.  
APPLICANT: Sven, Hastup  
TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose  
TITLE OF INVENTION: Or Hemicellulose  
NUMBER OF SEQUENCES: 85  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: No. 54570460 No. 5457046disk of No. 5457046th America, Inc.  
STREET: 405 Lexington Avenue, 62nd Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/361,920  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/940,860  
FILING DATE: 28-OCT-1992  
APPLICATION NUMBER: DK 1158/90  
FILING DATE: 09-MAY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK91/00124  
FILING DATE: 08-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 3435, 204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-867-0298  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 39 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
; US-08-361-920-32

Query Match 57.1%; Score 16; DB 1; Length 39;

Best Local Similarity 79.2%; Pred. No. 47;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 cctatgtgtgtccaaattgctc 25  
|| |||| | |||| |||| ||||  
Db 34 CCAATGTGTGCGCCAGATTGCTC 11

RESULT 2  
US-08-479-939-32/C  
; Sequence 32, Application US/08479939  
; Patent No. 5686593  
; GENERAL INFORMATION:  
; APPLICANT: Woelldike, Helle F.  
; APPLICANT: Hagen, Frederick  
; APPLICANT: Hjort, Carsten M.  
; APPLICANT: Sven, Hastrup  
; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose  
; NUMBER OF SEQUENCES: 85  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: No. 56865930 No. 5686593disk of No. 5686593th America, Inc.  
; STREET: 405 Lexington Avenue, 62nd Floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 10174-6201  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/479,939  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US/08/361,920  
; FILING DATE: 22-DEC-1994  
; APPLICATION NUMBER: US 07/940,860  
; FILING DATE: 28-OCT-1992  
; APPLICATION NUMBER: DK 1158/90  
; FILING DATE: 09-MAY-1990  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: PCT/DK91/00124  
; FILING DATE: 08-MAY-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lambiris, Elias J.  
; REGISTRATION NUMBER: 33,728  
; REFERENCE/DOCKET NUMBER: 3435.204-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-867-0298  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 39 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; US-08-479-939-32

Query Match 57.1%; Score 16; DB 1; Length 39;  
Best Local Similarity 79.2%; Pred. No. 47;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
OY 2 cctatgtgtgtccaaattgctc 25  
|| |||| | |||| |||| ||||  
Db 34 CCAATGTGTGCGCCAGATTGCTC 11

RESULT 3  
US-08-483-432-32/C  
; Sequence 32, Application US/08483432  
; Patent No. 5763254  
; GENERAL INFORMATION:  
; APPLICANT: Woelldike, Helle F.  
; APPLICANT: Hagen, Frederick  
; APPLICANT: Hjort, Carsten M.  
; APPLICANT: Sven, Hastrup  
; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose  
; NUMBER OF SEQUENCES: 85  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: No. 57632540 No. 5763254disk of No. 5763254th America, Inc.  
; STREET: 405 Lexington Avenue, 62nd Floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 10174-6201  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/483,432  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US/08/361,920  
; FILING DATE:  
; APPLICATION NUMBER: US 07/940,860  
; FILING DATE: 28-OCT-1992  
; APPLICATION NUMBER: DK 1158/90  
; FILING DATE: 09-MAY-1990  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: PCT/DK91/00124  
; FILING DATE: 08-MAY-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lambiris, Elias J.  
; REGISTRATION NUMBER: 33,728  
; REFERENCE/DOCKET NUMBER: 3435.204-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-867-0298  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 39 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; US-08-483-432-32

Query Match 57.1%; Score 16; DB 2; Length 39;  
Best Local Similarity 79.2%; Pred. No. 47;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
OY 2 cctatgtgtgtccaaattgctc 25  
|| |||| | |||| |||| ||||  
Db 34 CCAATGTGTGCGCCAGATTGCTC 11

```
RESULT 4
US-08-459-135A-2
; Sequence 2, Application US/08459135A
; Patent No. 595078
; GENERAL INFORMATION:
; APPLICANT: BURHAM, Martin Karl Russel
; APPLICANT: CHOIRA, Ian
; APPLICANT: CRITCHLEY, Ian Alfred
; APPLICANT: KNOWLES, David Justin Charles
; TITLE OF INVENTION: FIBROBLAST BINDING PROTEIN: MONOCLONAL ANTIBODY AND THEIR US
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fastseq, Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,135A
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB/94/00215
; FILING DATE: 04-Feb-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmil, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P30591C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: GENOMIC DNA
; US-08-459-135A-2

Query Match 53.6%; Score 15; DB 4; Length 34;
Best Local Similarity 78.3%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 cccatgttggtgccaattg 23
Db 7 CCTTACGTTGTCACGATTGG 29

RESULT 5
US-08-198-431-31/C
; Sequence 31, Application US/08198431
; Patent No. 560593
; GENERAL INFORMATION:
; APPLICANT: Stemmer, Willem P.C.
; TITLE OF INVENTION: Methods for In Vitro Recombination
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22313
```

```
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/198,431
; FILING DATE: 17-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mooi, Leslie
; REGISTRATION NUMBER: 37,047
; REFERENCE/DOCKET NUMBER: 000324-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-854-8275
; TELEFAX: 415-854-7400
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (oligonucleotide)
; US-08-198-431-31

Query Match 50.7%; Score 14.2; DB 1; Length 72;
Best Local Similarity 70.4%; Pred. No. 3.4e+02;
Matches 19; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 2 cctatgttggtgccaattg 28
Db 65 CTCTGTTGTACCAATATATGTACG 39

RESULT 6
US-08-564-955-31/C
; Sequence 31, Application US/08564955
; Patent No. 581128
; GENERAL INFORMATION:
; APPLICANT: STEMMER, WILLEM P.C.
; APPLICANT: CRAMER, ANDREAS M.
; TITLE OF INVENTION: METHODS FOR GENERATING POLYNUCLEOTIDES
; TITLE OF INVENTION: HAVING DESIRED CHARACTERISTICS BY ITERATIVE SELECTION AND
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW
; STREET: TWO EMBARCADERO CENTER, 8TH FLOOR
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: U.S.A.
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/564,955
; FILING DATE: 30-NOV-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/198,431
; FILING DATE: 17-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/537,874
; FILING DATE: 30-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02126
; FILING DATE: 17-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: DUNN, TRACY J.
```

```

?      REGISTRATION NUMBER: 34,587
?      REFERENCE/DOCKET NUMBER: 16528J-014611US
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: (415) 326-2400
?      TELEFAX: (415) 576-0300
?      INFORMATION FOR SEQ ID NO: 31:
?      SEQUENCE CHARACTERISTICS:
?          LENGTH: 72 base pairs
?          TYPE: nucleic acid
?          STRANDEDNESS: single
?          TOPOLOGY: linear
?
US-08-564-935-31

```

Query Match	50.7%	Score 14.2	DB 2	Length 72
Best Local Similarity	70.4%	Pred. No. 3.4e+02		
Matches 19	Conservative	0	Mismatches 8	Indels 0
Gaps				0
QY	2	cctatgctgtggtgccaaattggtcaag	28	
db	65	cttctggttggtaccatattatgtatgacg	39	

RESULT 7  
US-08-537-874-31/C  
: Sequence 31, Application US/08537874  
: Patent No. 5830721  
: GENERAL INFORMATION:  
: APPLICANT: Stemmer, Willem P.C.  
: APPLICANT: Cramerl, Andreas  
: TITLE OF INVENTION: DNA Mutagenesis by Random Fragmentation  
: TITLE OF INVENTION: and Reassembly  
: NUMBER OF SEQUENCES: 62  
: CORRESPONDENCE ADDRESS:  
:

ADDRESS: Townsend and Townsend and Crew LIFE  
 STREET: Two Embarcadero Center, 8th Floor  
 CITY: San Francisco  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94111  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.3.3  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/537,874  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: EP PCT/US95/02126  
 FILING DATE: 17-FEB-1995  
 APPLICATION NUMBER: US 08/198,431  
 FILING DATE: 17-FEB-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Liebeschuetz, Joe  
 REGISTRATION NUMBER: 37,505  
 REFERENCE/DOCKET NUMBER: 018097-014610  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-576-0200  
 TELEFAX: 415-576-0300  
 INFORMATION FOR SEO ID NO.: 31:  
 SOURCE CHARACTERISTICS:  
 LENGTH: 72 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IS-08-537-874-31

Query Match	50.7%;	Score 14.2;	DB 3;	length 72;
Best Local Similarity	70.4%;	Pred. No. 3.4e+02;		
Matches 19; Conservative	0;	Mismatches 8;	Indels 0;	Gaps 0;

Qy	2	c	c	a	t	a	t	t	g	t	g	t	g	c	a	a	a	a	a	t	t	g	t	c	a	a	g	28
Db	65	c	t	t	c	t	g	t	g	t	g	t	a	c	c	a	t	a	t	a	t	a	t	g	t	a	c	39

RESULT 8  
US-08-764-100-6  
Sequence 6, Application US/08764100  
Patent No. 5773700  
GENERAL INFORMATION:  
APPLICANT: van Grinsven J., Martinus Q.  
APPLICANT: De Haan, Petrus T.  
APPLICANT: Gielen L., Johannes J.  
APPLICANT: Peters, Dick  
APPLICANT: Goldbach, Robert W.  
TITLE OF INVENTION: Improvements in or Relating to Organic  
TITLE OF INVENTION: Compounds  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sandoz Agro, Inc  
STREET: 975 California Avenue  
CITY: Palo Alto

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,100
FILING DATE: 06-DEC-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,064
FILING DATE:
APPLICATION NUMBER: US 08/032,235
FILING DATE: 17-MAR-1993
APPLICATION NUMBER: GB 9206016.9
FILING DATE: 19-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: NO. 57737001s, Allen E.
REGISTRATION NUMBER: 34,490
REFERENCE/DOCKET NUMBER: 137-1061
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 354-3592
TELEFAX: (415) 857-1125
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 36 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-764-100-6

```

Query Match	50.08;	Score 14;	DB 2;	Length 36;
Best Local Similarity	77.38;	Pred. No. 3.7e+02;		
Matches 17;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;

Qy	4	tatgttggtgccaatgtgc	25
Db	14	TATATTCTGACACAATTGCTC	35

RESULT 9  
US-08-583-276-9  
; Sequence 9, Application US/08583276  
; Patent No. 5837536  
; GENERAL INFORMATION:  
; APPLICANT: McDonagh, Kevin T.



```

: APPLICANT: Nienhuis, Arthur
: APPLICANT: Tolstoshev, Paul
: TITLE OF INVENTION: IMPROVED EXPRESSION OF HUMAN
: TITLE OF INVENTION: MULTIDRUG RESISTANCE GENES AND IMPROVED
: TITLE OF INVENTION: SELECTION OF CELLS TRANSDUCED WITH SUCH GENES
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
: STREET: 6 Becker Farm Road
: CITY: Roseland
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07068
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch diskette
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: PC-DOS
: SOFTWARE: DNA.V2
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/583,276
: FILING DATE: 05-JAN-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/332,444
: FILING DATE: 31-OCT-1994
: APPLICATION NUMBER: 07/887,712
: FILING DATE: 22-MAY-1992
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 35 bases
: TYPE: nucleic acid
: STRANDEDNESS: singular
: TOPOLOGY: linear
: MOLECULE TYPE:
: DESCRIPTION: DNA primer
: US-08-583-276-9

```

```

Query Match          49.3%; Score 13.8; DB 3; Length 35;
Best Local Similarity 88.2%; Pred. No. 4.5e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 6 tttgtgtccaaatg 22
| | | | | | | | | | |
DB 19 TATTGTGCAAAATG 35

```

```

RESULT 10
: US-08-583-276-10/c
: Sequence 10, Application US/08583276
: Patent No. 5837536
: GENERAL INFORMATION:
: APPLICANT: McDonagh, Kevin T.
: APPLICANT: Nienhuis, Arthur
: APPLICANT: Tolstoshev, Paul
: TITLE OF INVENTION: IMPROVED EXPRESSION OF HUMAN
: TITLE OF INVENTION: MULTIDRUG RESISTANCE GENES AND IMPROVED
: TITLE OF INVENTION: SELECTION OF CELLS TRANSDUCED WITH SUCH GENES
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
: STREET: 6 Becker Farm Road
: CITY: Roseland
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07068
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch diskette
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: PC-DOS
: SOFTWARE: DNA.V2

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/583,276
: FILING DATE: 05-JAN-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/332,444
: FILING DATE: 31-OCT-1994
: APPLICATION NUMBER: 07/887,712
: FILING DATE: 22-MAY-1992
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 35 bases
: TYPE: nucleic acid
: STRANDEDNESS: singular
: TOPOLOGY: linear
: MOLECULE TYPE:
: DESCRIPTION: DNA primer
: US-08-583-276-10

```

```

Query Match          49.3%; Score 13.8; DB 3; Length 35;
Best Local Similarity 88.2%; Pred. No. 4.5e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 6 tttgtgtccaaatg 22
| | | | | | | | | | |
DB 17 TATTGTGCAAAATG 1

```

```

RESULT 11
: US-08-743-637B-210/c
: Sequence 210, Application US/08743637B
: Patent No. 5994066
: GENERAL INFORMATION:
: APPLICANT: BERGERON, Michel G.
: APPLICANT: PICARD, Francois J.
: APPLICANT: OUELLETTE, Marc
: APPLICANT: ROY, Paul H.
: TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
: TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
: TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
: TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
: NUMBER OF SEQUENCES: 273
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: QUARLES & BRADY
: STREET: 411 EAST WISCONSIN AVENUE
: CITY: MILWAUKEE
: STATE: WISCONSIN
: COUNTRY: USA
: ZIP: 53202-4497
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/743,637B
: FILING DATE: 04-NOV-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/526,840
: FILING DATE: 11-SEP-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: BAKER, Jean C.
: REGISTRATION NUMBER: 35,433
: REFERENCE/DOCKET NUMBER: 850586,90012
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (414) 277-5500
: TELEFAX: (414) 277-5591
: INFORMATION FOR SEQ ID NO: 210:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 21 base pairs
: TYPE: nucleic acid

```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-743-637B-210

Query Match          48.6%; Score 13.6; DB 4; Length 21;
Best Local Similarity 80.0%; Pred. No. 5.1e+02;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      7  gtgtgtgccaattgtcga 26
        ||||| ||||| |||||
Db      20  GTGCTGATTAAGTGTCA 1

RESULT 12
US-08-538-875-26
; Sequence 26, Application US/08538875
; Patent No. 5773582
; GENERAL INFORMATION:
; APPLICANT: Shin, Hang-Cheol
; APPLICANT: Shin, Nam-Kyu
; APPLICANT: Lee, Inkyung
; APPLICANT: Kang, Sungzong
; TITLE OF INVENTION: TUMOR NECROSIS FACTOR MUTEINS
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shin, Hang-Cheol
; STREET: Jukong Gocheng Apt. 1014-806, Haan-dong
; CITY: Kwangmyung-shi
; STATE: Kwangmyung-do
; COUNTRY: Republic of Korea
; ZIP: 423-060
; ADDRESSEE: Shin, Nam-Kyu
; STREET: #181-404 Sadang-4-dong, Dongjak-ku
; CITY: Seoul
; STATE: Seoul
; COUNTRY: Republic of Korea
; ZIP: 156-094
; ADDRESSEE: Lee, Inkyung
; STREET: 11/2, #302-39 Juan-4-dong, Nam-ku
; CITY: Incheon
; STATE: Incheon
; COUNTRY: Republic of Korea
; ZIP: 402-204
; ADDRESSEE: Kang, Sungzong
; STREET: #84-4 Daeshin-dong, Seodaemun-ku
; CITY: Seoul
; STATE: Seoul
; COUNTRY: Republic of Korea
; ZIP: 120-160
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette 3.5inch 2.0MB storage
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/538,875
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/193,336
; FILING DATE:
; APPLICATION NUMBER: KR 93-1751
; FILING DATE: 9-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:
; TELEFAX:
; INFORMATION FOR SEQ ID NO: 26:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: primer DNA
US-08-538-875-26

Query Match          48.6%; Score 13.6; DB 2; Length 26;
Best Local Similarity 80.0%; Pred. No. 5.3e+02;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      6  tgtgtgtgccaattgtc 25
        || ||||| ||||| ||
Db      1  TGGTGTGTCACAAAAGGCG 20

RESULT 13
US-09-289-376-21/c
; Sequence 21, Application US/09289376
; Patent No. 6013788
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; APPLICANT: Lex M. Cowsett
; TITLE OF INVENTION: ANTISENSE MODULATION OF SMAD3 EXPRESSION
; FILE REFERENCE: RTS-0043
; CURRENT APPLICATION NUMBER: US/09/289,376
; CURRENT FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 21
; LENGTH: 18
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Antisense Oligonucleotide
US-09-289-376-21

Query Match          47.9%; Score 13.4; DB 5; Length 18;
Best Local Similarity 93.3%; Pred. No. 6e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2  cctatgtgtgtgcca 16
        ||| ||||| |||||
Db      18  CCGTGTGTGCGCA 4

RESULT 14
US-08-592-406-5
; Sequence 5, Application US/08592406
; Patent No. 5821059
; GENERAL INFORMATION:
; APPLICANT: MINION, F. CHRIS
; APPLICANT: KNOTTSON, Kevin L.
; TITLE OF INVENTION: MYCOPLASMA EXPRESSION SYSTEM
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,406
; FILING DATE: 06-FEB-1996
; CLASSIFICATION: 435
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:05:32 ; Search time 1236.38 Seconds

(without alignments)  
-22.031 Million cell updates/sec

Title: US-09-164-714-9

Perfect score: 28  
Sequence: 1 agatgcacgaacatcaacgtaagaac 28

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 356616

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_om:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vl:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_or:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vl:\*  
32: gb\_htg1:\*  
33: gb\_htg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pr4:\*  
41: gb\_htg3:\*  
42: gb\_htg4:\*  
43: gb\_htg5:\*  
44: gb\_htg6:\*

45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pr5:\*  
52: gb\_htg8:\*  
53: gb\_htg9:\*  
54: gb\_htg10:\*  
55: gb\_htg11:\*  
56: gb\_htg12:\*  
57: gb\_htg13:\*  
58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	16	57.1	79 5 107076	107076 Sequence 2
2	16	57.1	79 5 183771	183771 Sequence 8
3	15.8	56.4	54 5 A57944	A57944 Sequence 10
4	15.8	56.4	54 5 A57945	A57945 Sequence 11
5	15.8	56.4	54 5 A57946	A57946 Sequence 12
6	15.8	56.4	54 5 A57947	A57947 Sequence 13
7	15.8	56.4	54 5 A57948	A57948 Sequence 14
8	15.8	56.4	54 5 A57949	A57949 Sequence 15
9	15.8	56.4	54 5 A57950	A57950 Sequence 16
10	15.8	56.4	54 5 A57951	A57951 Sequence 17
11	15.8	56.4	54 5 A57952	A57952 Sequence 18
12	15.8	56.4	54 5 A57954	A57954 Sequence 20
13	15.8	56.4	54 5 A57956	A57956 Sequence 22
14	15.8	56.4	54 5 A57957	A57957 Sequence 23
15	15.8	56.4	54 5 A57959	A57959 Sequence 25
16	15.8	56.4	54 5 A57960	A57960 Sequence 26
17	15.8	56.4	54 5 A57961	A57961 Sequence 27
18	15.8	56.4	92 5 E05823	E05823 DNA encodin
19	15	53.6	31 5 I37030	I37030 Sequence 43
20	15	53.6	31 5 I93880	I93880 Sequence 43
21	14.8	52.9	53 5 AR053523	AR053523 Sequence
22	14.8	52.9	54 5 A57955	A57955 Sequence 21
23	14.8	52.9	54 5 A57958	A57958 Sequence 24
24	14.8	52.9	54 5 A57965	A57965 Sequence 31
25	14.6	52.1	43 5 A46553	A46553 Sequence 9
26	14.6	52.1	69 5 AR035210	AR035210 Sequence
27	14.6	52.1	69 5 AR035215	AR035215 Sequence
28	14.2	50.7	47 5 A84021	A84021 Sequence 8
29	14.2	50.7	54 5 A57943	A57943 Sequence 9
30	14.2	50.7	54 5 A57953	A57953 Sequence 19
31	14.2	50.7	74 12 U52651	U52651 Rattus norv
32	14.2	50.7	79 9 HDGCC3XC	HDGCC3XC
33	14.2	50.7	99 5 AR008442	AR008442 Sequence
34	14.2	50.7	99 5 I87434	I87434 Sequence 69
35	13.8	49.3	23 5 A44315	A44315 Sequence 1
36	13.8	49.3	30 5 A44320	A44320 Sequence 6
37	13.8	49.3	42 5 AR026320	AR026320 Sequence
38	13.8	49.3	98 10 HS292873	HS292873 Homo sapien
39	13.6	48.6	21 5 I76941	I76941 Sequence 7
40	13.6	48.6	23 5 I37347	I37347 Sequence 36
41	13.6	48.6	23 5 I94197	I94197 Sequence 36
42	13.6	48.6	29 5 I83408	I83408 Sequence 9
43	13.6	48.6	32 5 AR035783	AR035783 Sequence
44	13.6	48.6	35 5 A79354	A79354 Sequence 3
45	13.6	48.6	36 5 AR001579	AR001579 Sequence

#### ALIGNMENTS

RESULT 1  
LOCUS 107076 79 bp PAT 02-DEC-1994  
DEFINITION Sequence 2 from Patent EP 03144415.  
ACCESSION 107076  
VERSION 107076.1 GI:590362  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 79)  
AUTHORS Namen,A.E., Goodwin,R.G., Lupton,S.D., Mochizuki,D.Y., Price,V.L.  
and Deeley,M.C.  
TITLE Interleukin-7  
JOURNAL Patent: EP 0314415-A2 2 03-MAY-1989;  
FEATURES Location/Qualifiers  
source 1..79  
BASE COUNT 27 a 18 c 22 g 12 t  
ORIGIN

Query Match 57.1%; Score 16; DB 5; Length 79;  
Best Local Similarity 79.2%; Pred. No. 5.8e+03;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 gatgccaaagaatcaacgtaag 25  
Db 34 GATGACAGCAAAATCGAAGTAGG 57

RESULT 2  
LOCUS 183771 79 bp DNA PAT 10-AUG-1998  
DEFINITION Sequence 8 from patent US 5714585.  
ACCESSION 183771  
VERSION 183771.1 GI:3407301  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 79)  
AUTHORS Namen,A.E., Goodwin,R.G., Lupton,S.D. and Mochizuki,D.Y.  
TITLE Antibodies that are immunoreactive with Interleukin-7  
JOURNAL Patent: US 5714585-A 8 03-FEB-1998;  
FEATURES Location/Qualifiers  
source 1..79  
BASE COUNT 27 a 18 c 22 g 12 t  
ORIGIN

Query Match 57.1%; Score 16; DB 5; Length 79;  
Best Local Similarity 79.2%; Pred. No. 5.8e+03;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 gatgccaaagaatcaacgtaag 25  
Db 34 GATGACAGCAAAATCGAAGTAGG 57

RESULT 3  
LOCUS A57944 54 bp DNA PAT 05-MAR-1998  
DEFINITION Sequence 10 from Patent EP0743364.  
ACCESSION A57944  
VERSION A57944.1 GI:3713714  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 54)

AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 10 20-NOV-1996;  
COMMENT COMMISSARIAT ENERGIE ATOMIQUE (FR)  
Other publication FR 2734281 961122.  
FEATURES Location/Qualifiers  
source 1..54  
BASE COUNT 31 a 4 c 13 g 6 t  
ORIGIN

Query Match 56.4%; Score 15.8; DB 5; Length 54;  
Best Local Similarity 74.1%; Pred. No. 7.2e+03;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agatgccaaagaatcaacgtaaga 27  
Db 21 AGAGGAAGACCAAAACAAAGTAAGAA 47

RESULT 4  
LOCUS A57945 54 bp DNA PAT 05-MAR-1998  
DEFINITION Sequence 11 from Patent EP0743364.  
ACCESSION A57945  
VERSION A57945.1 GI:3713715  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 11 20-NOV-1996;  
COMMENT COMMISSARIAT ENERGIE ATOMIQUE (FR)  
Other publication FR 2734281 961122.  
FEATURES Location/Qualifiers  
source 1..54  
BASE COUNT 30 a 4 c 15 g 5 t  
ORIGIN

Query Match 56.4%; Score 15.8; DB 5; Length 54;  
Best Local Similarity 74.1%; Pred. No. 7.2e+03;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agatgccaaagaatcaacgtaaga 27  
Db 21 AGAGGAAGACCAAAACAAAGTAAGAA 47

RESULT 5  
LOCUS A57946 54 bp DNA PAT 05-MAR-1998  
DEFINITION Sequence 12 from Patent EP0743364.  
ACCESSION A57946  
VERSION A57946.1 GI:3713716  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of

HIV-1 mother-foetal transmission  
Patent: EP 0743364-A 12-20-NOV-1996;  
COMMISSARIAT ENERGIE ATOMIQUE (FR)  
Other publication FR 2734281 961122.  
Location/Qualifiers

COMMENT  
source  
1. .54  
/db\_xref="taxon:32644"

BASE COUNT 31 a 4 c 14 g 5 t  
ORIGIN

Query Match 56.4%; Score 15.8; DB 5; Length 54;  
Best Local Similarity 74.1%; Pred. No. 7.2e+03;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agatgccaaagcaatcaacgtaagaa 27  
||| | ||||| ||| |||||  
Db 21 AGAGGAAGAGCAAAACAAAGTAAGAA 47

RESULT 6  
LOCUS A57947 54 bp DNA PAT 05-MAR-1998  
DEFINITION Sequence 13 from Patent EP0743364.  
ACCESSION A57947  
VERSION A57947.1 GI:3713717  
KEYWORDS  
SOURCE .  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 13-20-NOV-1996;  
COMMENT COMMISSARIAT ENERGIE ATOMIQUE (FR)  
FEATURES Other publication FR 2734281 961122.  
source Location/Qualifiers  
1. .54  
/db\_xref="taxon:32644"

BASE COUNT 31 a 4 c 13 g 6 t  
ORIGIN

Query Match 56.4%; Score 15.8; DB 5; Length 54;  
Best Local Similarity 74.1%; Pred. No. 7.2e+03;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agatgccaaagcaatcaacgtaagaa 27  
||| | ||||| ||| |||||  
Db 21 AGAGGAAGAGCAAAACAAAGTAAGAA 47

RESULT 7  
LOCUS A57948 54 bp DNA PAT 05-MAR-1998  
DEFINITION Sequence 14 from Patent EP0743364.  
ACCESSION A57948  
VERSION A57948.1 GI:3713718  
KEYWORDS  
SOURCE .  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 14-20-NOV-1996;  
COMMENT COMMISSARIAT ENERGIE ATOMIQUE (FR)

Other publication FR 2734281 961122.  
Location/Qualifiers

COMMENT  
source  
1. .54  
/db\_xref="taxon:32644"

BASE COUNT 32 a 5 c 12 g 5 t  
ORIGIN

Query Match 56.4%; Score 15.8; DB 5; Length 54;  
Best Local Similarity 74.1%; Pred. No. 7.2e+03;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agatgccaaagcaatcaacgtaagaa 27  
||| | ||||| ||| |||||  
Db 21 AGAGGAAGAGCAAAACAAAGTAAGAA 47

RESULT 8  
LOCUS A57949 54 bp DNA PAT 05-MAR-1998  
DEFINITION Sequence 15 from Patent EP0743364.  
ACCESSION A57949  
VERSION A57949.1 GI:3713719  
KEYWORDS  
SOURCE .  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 15-20-NOV-1996;  
COMMENT COMMISSARIAT ENERGIE ATOMIQUE (FR)  
FEATURES Other publication FR 2734281 961122.  
source Location/Qualifiers  
1. .54  
/db\_xref="taxon:32644"

BASE COUNT 30 a 5 c 14 g 5 t  
ORIGIN

Query Match 56.4%; Score 15.8; DB 5; Length 54;  
Best Local Similarity 74.1%; Pred. No. 7.2e+03;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agatgccaaagcaatcaacgtaagaa 27  
||| | ||||| ||| |||||  
Db 21 AGAGGAAGAGCAAAACAAAGTAAGAA 47

RESULT 9  
LOCUS A57950 54 bp DNA PAT 05-MAR-1998  
DEFINITION Sequence 16 from Patent EP0743364.  
ACCESSION A57950  
VERSION A57950.1 GI:3713720  
KEYWORDS  
SOURCE .  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 16-20-NOV-1996;  
COMMENT COMMISSARIAT ENERGIE ATOMIQUE (FR)  
FEATURES Other publication FR 2734281 961122.  
source Location/Qualifiers  
1. .54

BASE COUNT 30 a 5 c 14 g 5 t  
ORIGIN /organism="unidentified"  
/db\_xref="taxon:32644"

Query Match 56.4%; Score 15.8; DB 5; Length 54;  
Best Local Similarity 74.1%; Pred. No. 7.2e+03;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agatgccaaagcaataacgtaagaa 27  
||| | ||||| ||| ||||| |||  
Db 21 AGAGGAAGACCAAAAGTAAGAA 47

RESULT 10 A57951 54 bp DNA PAT 05-MAR-1998  
LOCUS A57951  
DEFINITION Sequence 17 from Patent EP0743364.  
ACCESSION A57951  
VERSION A57951.1 GI:3713721  
KEYWORDS  
SOURCE .  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narva,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding  
fragments and their application as reactives for risk evaluation of  
HIV-1 mother-foetal transmission

JOURNAL Patent: EP 0743364-A 17 20-NOV-1996;  
COMMISSARIAT ENERGIE ATOMIQUE (FR)  
Other publication FR 2734281 961122.  
COMMENT Location/Qualifiers  
FEATURES  
source 1..54  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 31 a 4 c 14 g 5 t  
ORIGIN /db\_xref="taxon:32644"

Query Match 56.4%; Score 15.8; DB 5; Length 54;  
Best Local Similarity 74.1%; Pred. No. 7.2e+03;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agatgccaaagcaataacgtaagaa 27  
||| | ||||| ||| ||||| |||  
Db 21 AGAGGAAGACCAAAAGTAAGAA 47

RESULT 11 A57952 54 bp DNA PAT 05-MAR-1998  
LOCUS A57952  
DEFINITION Sequence 18 from Patent EP0743364.  
ACCESSION A57952  
VERSION A57952.1 GI:3713722  
KEYWORDS  
SOURCE .  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narva,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding  
fragments and their application as reactives for risk evaluation of  
HIV-1 mother-foetal transmission

JOURNAL Patent: EP 0743364-A 18 20-NOV-1996;  
COMMISSARIAT ENERGIE ATOMIQUE (FR)  
Other publication FR 2734281 961122.  
COMMENT Location/Qualifiers  
FEATURES  
source 1..54  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 29 a 5 c 15 g 4 t 1 others

ORIGIN

Query Match 56.4%; Score 15.8; DB 5; Length 54;  
Best Local Similarity 74.1%; Pred. No. 7.2e+03;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agatgccaaagcaataacgtaagaa 27  
||| | ||||| ||| ||||| |||  
Db 21 AGAGGAAGACCAAAAGTAAGAA 47

RESULT 12 A57954 54 bp DNA PAT 05-MAR-1998  
LOCUS A57954  
DEFINITION Sequence 20 from Patent EP0743364.  
ACCESSION A57954  
VERSION A57954.1 GI:3713724  
KEYWORDS  
SOURCE .  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narva,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding  
fragments and their application as reactives for risk evaluation of  
HIV-1 mother-foetal transmission

JOURNAL Patent: EP 0743364-A 20 20-NOV-1996;  
COMMISSARIAT ENERGIE ATOMIQUE (FR)  
Other publication FR 2734281 961122.  
COMMENT Location/Qualifiers  
FEATURES  
source 1..54  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 30 a 4 c 15 g 5 t  
ORIGIN /db\_xref="taxon:32644"

Query Match 56.4%; Score 15.8; DB 5; Length 54;  
Best Local Similarity 74.1%; Pred. No. 7.2e+03;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agatgccaaagcaataacgtaagaa 27  
||| | ||||| ||| ||||| |||  
Db 21 AGAGGAAGACCAAAAGTAAGAA 47

RESULT 13 A57956 54 bp DNA PAT 05-MAR-1998  
LOCUS A57956  
DEFINITION Sequence 22 from Patent EP0743364.  
ACCESSION A57956  
VERSION A57956.1 GI:3713726  
KEYWORDS  
SOURCE .  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narva,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding  
fragments and their application as reactives for risk evaluation of  
HIV-1 mother-foetal transmission

JOURNAL Patent: EP 0743364-A 22 20-NOV-1996;  
COMMISSARIAT ENERGIE ATOMIQUE (FR)  
Other publication FR 2734281 961122.  
COMMENT Location/Qualifiers  
FEATURES  
source 1..54  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 30 a 6 c 14 g 4 t





---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:24:06 ; Search time 322.35 Seconds  
(without alignments)  
21.732 Million cell updates/sec

Title: US-09-164-714-9

Sequence: 1 agatgccaaagcaatcaacgtaagaac 28

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 431286

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15.8	56.4	54	1	T43658 HIV-1 matrix prote
2	15.8	56.4	54	1	T43659 HIV-1 matrix prote
3	15.8	56.4	54	1	T43662 HIV-1 matrix prote
4	15.8	56.4	54	1	T43643 HIV-1 matrix prote
5	15.8	56.4	54	1	T43644 HIV-1 matrix prote
6	15.8	56.4	54	1	T43645 HIV-1 matrix prote
7	15.8	56.4	54	1	T43646 HIV-1 matrix prote
8	15.8	56.4	54	1	T43647 HIV-1 matrix prote
9	15.8	56.4	54	1	T43648 HIV-1 matrix prote
10	15.8	56.4	54	1	T43649 HIV-1 matrix prote
11	15.8	56.4	54	1	T43650 HIV-1 matrix prote
12	15.8	56.4	54	1	T43652 HIV-1 matrix prote
13	15.8	56.4	54	1	T43654 HIV-1 matrix prote
14	15.8	56.4	54	1	T43655 HIV-1 matrix prote
15	15.8	56.4	54	1	T43657 HIV-1 matrix prote
16	15.8	56.4	92	1	Q53311 Hybrid antigen pro
17	15.2	54.3	38	1	N94509 Probe for N-termin
18	15.2	54.3	38	1	V83523 PCR primer used to
19	15.2	54.3	38	1	V83525 PCR primer used to
20	15.2	54.3	97	1	T19416 Human gene signatu
21	15.2	53.6	31	1	Q93504 Human stromelysin
22	14.8	52.9	47	1	X52584 Human genome biall
23	14.8	52.9	47	1	X52596 Human genome biall
24	14.8	52.9	47	1	X52546 Human genome biall
25	14.8	52.9	47	1	X52544 Human genome biall
26	14.8	52.9	53	1	V68220 PCR primer used to
27	14.8	52.9	54	1	O86781 SSI-1 mimetic prim
28	14.8	52.9	54	1	Q74004 SSI-1 mimetic PCR
29	14.8	52.9	54	1	T43663 HIV-1 matrix prote
30	14.8	52.9	54	1	T43663 HIV-1 matrix prote
31	14.6	52.1	54	1	T43656 HIV-1 matrix prote
32	14.6	52.1	69	1	T01299 PCR primer oligo D
33	14.6	52.1	69	1	Q30897 Primer 312-69. New
34	14.6	52.1	69	1	Q30892 Primer 312-64. New

C 35	14.6	52.1	100	1	T42735	Primer J089 for hu
C 36	14.4	51.4	60	1	T62581	PCR primer for cys
C 37	14.2	50.7	47	1	V68272	Penicillium chryso
C 38	14.2	50.7	54	1	T43651	HIV-1 matrix prote
C 39	14.2	50.7	54	1	T43641	HIV-1 matrix prote
C 40	14.2	50.7	99	1	V10203	Stealth virus nucl
C 41	14.2	50.7	99	1	V12016	Stealth virus plas
C 42	14	50.0	99	1	T30883	Primer 13 for 95 k
C 43	13.8	49.3	23	1	O88336	PCR primer OTG3042
C 44	13.8	49.3	30	1	O88341	PCR primer OTG5988
C 45	13.8	49.3	37	1	V39939	Streptococcus pneu

## ALIGNMENTS

RESULT 1	
T43658	
ID T43658 standard; DNA: 54 BP.	
AC T43658;	
DT 19-AUG-1997 (first entry)	
DE HIV-1 matrix protein p17 gene fragment 2754.	
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;	
KW detection; maternal transmission; hybridisation assay; immunoassay; ss.	
OS Human immunodeficiency virus type 1.	
PN Ep-743364-A2.	
PD 20-NOV-1996.	
PE 17-MAY-1996; 401084.	
PR 18-MAY-1995; FR-005914.	
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.	
PI Narwa R, Rogues P;	
DR WPI: 96-507733/51.	
PT Human immunodeficiency virus p17 gene fragments, derived proteins	
PT and antibodies - useful for assessing the risk of maternal	
PT transmission of HIV-1 infection	
PS Claim 3; Page 27; 46pp. French.	
CC This sequence is a specifically claimed example of 21-90 nucleotide	
CC long nucleic acid fragments, derived from the gene encoding part of the	
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at	
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;	
CC Y6 = ATA, TTA, CTG, GTA, CTA, GTC or ATC; Y7 = GAG or GAA; Y8 = GAA or	
CC AAG; Y9 = ATA, GTA, CTA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or	
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5	
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,	
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.	
CC The new nucleic acid fragments can be used as reagents for determining	
CC and assessing the risk of maternal-foetal transmission of HIV-1, using	
CC standard hybridisation or immuno assays. The presence of such sequences	
CC in maternal blood is strongly correlated with transmission of infection.	
SO Sequence 54 BP; 30 A; 4 C; 15 G; 5 T;	
Query Match 56.4%; Score 15.8; DB 1; Length 54;	
Best Local Similarity 74.1%; Pred. No. 2.1e+02;	
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;	
QY 1 agatgccaaagcaatcaacgtaagaac 27	
DB 21 AGAGCAAGAGCAACAAACAAAGTAACAA 47	
RESULT 2	
T43659	
ID T43659 standard; DNA: 54 BP.	
AC T43659;	
DT 19-AUG-1997 (first entry)	
DE HIV-1 matrix protein p17 gene fragment 2826.	
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;	
KW detection; maternal transmission; hybridisation assay; immunoassay; ss.	
OS Human immunodeficiency virus type 1.	
PN Ep-743364-A2.	
PD 20-NOV-1996.	
PE 17-MAY-1996; 401084.	

PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R. Roques P;  
DR WPI; 96-507733/51.  
PT Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal  
PT transmission of HIV-1 infection  
PS Claim 3; Page 27; 46pp; French.  
CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-foetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 30 A; 4 C; 15 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 agatgccaaagcaaatcaacgtaagaa 27  
||| | ||||| ||| ||||| |||  
Db 21 AGAGGAGAGCAAAACAAAGTAAGAA 47

RESULT 3  
T43642  
ID T43642 standard; DNA; 54 BP.  
AC T43642;  
DE 19-AUG-1997 (first entry)  
DT HIV-1 matrix protein p17 gene fragment ARI.  
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
KM detection; maternal transmission; hybridisation assay; immunoassay; ss.  
OS Human immunodeficiency virus type 1.  
PN EP-743364-A2.  
PD 20-NOV-1996.  
PF 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R. Roques P;  
DR WPI; 96-507733/51.  
PT Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal  
PT transmission of HIV-1 infection  
PS Claim 3; Page 23; 46pp; French.  
CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-foetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 31 A; 4 C; 13 G; 6 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 agatgccaaagcaaatcaacgtaagaa 27  
||| | ||||| ||| ||||| |||  
Db 21 AGAGGAGAGCAAAACAAAGTAAGAA 47

RESULT 4  
T43643  
ID T43643 standard; DNA; 54 BP.  
AC T43643;  
DE 19-AUG-1997 (first entry)  
DT HIV-1 matrix protein p17 gene fragment BOI.  
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
KM detection; maternal transmission; hybridisation assay; immunoassay; ss.  
OS Human immunodeficiency virus type 1.  
FH Key Location/Qualifiers  
FT mat\_peptide 1..54  
FT /tag= a  
FT /product= BOI\_peptide  
PN EP-743364-A2.  
PD 20-NOV-1996.  
PF 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R. Roques P;  
DR WPI; 96-507733/51.  
DR P-PSDB; W06610.  
PT Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal  
PT transmission of HIV-1 infection  
PS Claim 3; Page 23; 46pp; French.  
CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-foetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 30 A; 4 C; 15 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 agatgccaaagcaaatcaacgtaagaa 27  
||| | ||||| ||| ||||| |||  
Db 21 AGAGGAGAGCAAAACAAAGTAAGAA 47

RESULT 5  
T43644  
ID T43644 standard; DNA; 54 BP.  
AC T43644;  
DE 19-AUG-1997 (first entry)  
DT HIV-1 matrix protein p17 gene fragment DMU.  
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
KM detection; maternal transmission; hybridisation assay; immunoassay; ss.  
OS Human immunodeficiency virus type 1.  
PN EP-743364-A2.  
PD 20-NOV-1996.  
PF 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R. Roques P;  
DR WPI; 96-507733/51.  
PT Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal

PT transmission of HIV-1 infection  
PS Claim 3: Page 24; 46pp: French.  
CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, CTA, GTA, GTC or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CCA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-fetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 31 A; 4 C; 14 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 agatgccaaagcaatcaacgyltaagaa 27  
||| | ||||| ||| |||||  
Db 21 AGAGCAAGACGCAAAACAAAGTAAAGAA 47

RESULT 6  
T43645  
ID T43645 standard; DNA; 54 BP.  
AC T43645;  
DE 19-AUG-1997 (first entry)  
DE HIV-1 matrix protein p17 gene fragment PAT.  
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
KW detection; maternal transmission; hybridisation assay; immunoassay; ss.  
OS Human immunodeficiency virus type 1.  
PN EP-743364-A2.  
PD 20-NOV-1996.  
PF 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R, Roques P;  
PI WPI; 96-507733/51.  
DR Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal  
PT transmission of HIV-1 infection  
PS Claim 3: Page 24; 46pp: French.  
CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, CTA, GTA, GTC or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CCA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-fetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 31 A; 4 C; 13 G; 6 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 agatgccaaagcaatcaacgyltaagaa 27  
||| | ||||| ||| |||||  
Db 21 AGAGCAAGACGCAAAACAAAGTAAAGAA 47

RESULT 7

T43646  
ID T43646 standard; DNA; 54 BP.  
AC T43646;  
DE 19-AUG-1997 (first entry)  
DE HIV-1 matrix protein p17 gene fragment RYO.  
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
KW detection; maternal transmission; hybridisation assay; immunoassay; ss.  
OS Human immunodeficiency virus type 1.  
PN EP-743364-A2.  
PD 20-NOV-1996.  
PF 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R, Roques P;  
PI WPI; 96-507733/51.  
DR Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal  
PT transmission of HIV-1 infection  
PS Claim 3: Page 24; 46pp: French.  
CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, CTA, GTA, GTC or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CCA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
SQ Sequence 54 BP; 32 A; 5 C; 12 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 agatgccaaagcaatcaacgyltaagaa 27  
||| | ||||| ||| |||||  
Db 21 AGAGCAAGACGCAAAACAAAGTAAAGAA 47

RESULT 8  
T43647  
ID T43647 standard; DNA; 54 BP.  
AC T43647;  
DE 19-AUG-1997 (first entry)  
DE HIV-1 matrix protein p17 gene fragment FLO.  
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
KW detection; maternal transmission; hybridisation assay; immunoassay; ss.  
OS Human immunodeficiency virus type 1.  
PN EP-743364-A2.  
PD 20-NOV-1996.  
PF 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R, Roques P;  
PI WPI; 96-507733/51.  
DR Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal  
PT transmission of HIV-1 infection  
PS Claim 3: Page 24; 46pp: French.  
CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, CTA, GTA, GTC or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CCA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.

CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-foetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 30 A; 5 C; 14 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 agatgccaaagcaatcaacgtaagaa 27  
||| | ||||| ||| ||||| |||  
Db 21 AGAGGAAGACCAAAACAAAGTAAGAA 47

RESULT 9  
T43648  
ID T43648 standard; DNA; 54 BP.

AC T43648; DT 19-AUG-1997 (first entry)

DE HIV-1 matrix protein p17 gene fragment 4501.

KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
CC detection; maternal transmission; hybridisation assay; immunoassay; ss.

OS Human immunodeficiency virus type 1.  
FH Key location/Qualifiers  
FT mat\_peptide 1..54  
FT /\*tag= a

PN EP-743364-A2. /product= 4501\_peptide

PD 20-NOV-1996.

PF 17-MAY-1996; 401084.

PR 18-MAY-1995; FR-005914.

PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.

PI Narwa R, Roques P;

DR WPI: 96-507733/51.

PT P-PSDB; W06611

PT Human immunodeficiency virus p17 gene fragments, derived proteins

PT and antibodies - useful for assessing the risk of maternal

PT transmission of HIV-1 infection

PS Claim 3; Page 24; 46pp; French.

CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at

CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;

CC Y6 = ATA, TTA, CTG, CTA, GTG or ATG; Y7 = GAG or GAA; Y8 = GAA or

CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or

CC CGA; Y11 = AAA, AAT, CAT, AAC or ACC; provided that when Y5

CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,

CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.

CC The new nucleic acid fragments can be used as reagents for determining

CC and assessing the risk of maternal-foetal transmission of HIV-1, using

CC standard hybridisation or immuno assays. The presence of such sequences

CC in maternal blood is strongly correlated with transmission of infection.

CC Sequence 54 BP; 30 A; 5 C; 14 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 agatgccaaagcaatcaacgtaagaa 27  
||| | ||||| ||| ||||| |||  
Db 21 AGAGGAAGACCAAAACAAAGTAAGAA 47

RESULT 10

T43649

ID T43649 standard; DNA; 54 BP.

AC T43649; DT 19-AUG-1997 (first entry)

DE HIV-1 matrix protein p17 gene fragment FAI.

KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;

KW detection; maternal transmission; hybridisation assay; immunoassay; ss.  
OS Human immunodeficiency virus type 1.

PN EP-743364-A2.

PD 20-NOV-1996.

PF 17-MAY-1996; 401084.

PR 18-MAY-1995; FR-005914.

PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.

PI Narwa R, Roques P;

DR WPI: 96-507733/51.

PT Human immunodeficiency virus p17 gene fragments, derived proteins

PT and antibodies - useful for assessing the risk of maternal

PT transmission of HIV-1 infection

PS Claim 3; Page 25; 46pp; French.

CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at

CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;

CC Y6 = ATA, TTA, CTG, CTA, GTG or ATG; Y7 = GAG or GAA; Y8 = GAA or

CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or

CC CGA; Y11 = AAA, AAT, CAT, AAC or ACC; provided that when Y5

CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,

CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.

CC The new nucleic acid fragments can be used as reagents for determining

CC and assessing the risk of maternal-foetal transmission of HIV-1, using

CC standard hybridisation or immuno assays. The presence of such sequences

CC in maternal blood is strongly correlated with transmission of infection.

CC Sequence 54 BP; 31 A; 4 C; 14 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 agatgccaaagcaatcaacgtaagaa 27  
||| | ||||| ||| ||||| |||  
Db 21 AGAGGAAGACCAAAACAAAGTAAGAA 47

RESULT 11  
T43650  
ID T43650 standard; DNA; 54 BP.

AC T43650; DT 19-AUG-1997 (first entry)

DE HIV-1 matrix protein p17 gene fragment HAR.

KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
CC detection; maternal transmission; hybridisation assay; immunoassay; ss.

OS Human immunodeficiency virus type 1.

FH Key location/Qualifiers

FT mat\_peptide 1..54

FT /\*tag= a

PN EP-743364-A2. /product= HAR\_peptide

PD 20-NOV-1996.

PF 17-MAY-1996; 401084.

PR 18-MAY-1995; FR-005914.

PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.

PI Narwa R, Roques P;

DR WPI: 96-507733/51.

PT P-PSDB; W06612.

PT Human immunodeficiency virus p17 gene fragments, derived proteins

PT and antibodies - useful for assessing the risk of maternal

PT transmission of HIV-1 infection

PS Claim 3; Page 25; 46pp; French.

CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at

CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;

CC Y6 = ATA, TTA, CTG, CTA, GTG or ATG; Y7 = GAG or GAA; Y8 = GAA or

CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or

CC CGA; Y11 = AAA, AAT, CAT, AAC or ACC; provided that when Y5

CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,

CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.

CC The new nucleic acid fragments can be used as reagents for determining

CC and assessing the risk of maternal-foetal transmission of HIV-1, using

CC standard hybridisation or immuno assays. The presence of such sequences

CC in maternal blood is strongly correlated with transmission of infection.

CC and assessing the risk of maternal-foetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP: 29 A; 5 C; 15 G; 4 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 agatgccaaagcaatcaacgtaagaa 27  
||| | ||||| ||| |||||  
Db 21 AGAGGAGAGCAAAACAAAGTAAGAA 47

RESULT 12  
T43652  
ID T43652 standard; DNA: 54 BP.  
AC T43652;  
DE 19-AUG-1997 (first entry)  
DT HIV-1 matrix protein p17 gene fragment CHET.  
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
OS Human immunodeficiency virus type 1.  
PN EP-743364-A2.  
PF 20-NOV-1996.  
PR 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R, Roques P;  
DR WPI: 96-507733/51.

PT Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal

PT transmission of HIV-1 infection  
PS Claim 3; Page 26; 46pp; French.

CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, CTA, GTG or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-foetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP: 30 A; 4 C; 15 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;

Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 agatgccaaagcaatcaacgtaagaa 27  
||| | ||||| ||| |||||  
Db 21 AGAGGAGAGCAAAACAAAGTAAGAA 47

RESULT 13  
T43654  
ID T43654 standard; DNA: 54 BP.

AC T43654;  
DE 19-AUG-1997 (first entry)

DT HIV-1 matrix protein p17 gene fragment SIW.

KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;

OS Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
OS Human immunodeficiency virus type 1.

PN Key Location/Qualifiers

FT mat\_peptide 1..54

FT /\*tag- a

FT /product- SIW\_peptide

FT

PN EP-743364-A2.

PD 20-NOV-1996.

PF 17-MAY-1996; 401084.

PR 18-MAY-1995; FR-005914.

PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.

PI Narwa R, Roques P;

DR WPI: 96-507733/51.

DR P-PSDB: W06614.

PT Human immunodeficiency virus p17 gene fragments, derived proteins

PT and antibodies - useful for assessing the risk of maternal

PT transmission of HIV-1 infection  
PS Claim 3; Page 26; 46pp; French.

CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, CTA, GTG or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-foetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP: 30 A; 6 C; 14 G; 4 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 agatgccaaagcaatcaacgtaagaa 27  
||| | ||||| ||| |||||  
Db 21 AGAGGAGAGCAAAACAAAGTAAGAA 47

RESULT 14

T43655  
ID T43655 standard; DNA: 54 BP.

AC T43655;

DE 19-AUG-1997 (first entry)

DT HIV-1 matrix protein p17 gene fragment MOE.

KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;

OS Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
OS Human immunodeficiency virus type 1.

PN Key Location/Qualifiers

FT mat\_peptide 1..54

FT /\*tag- a

FT /product- MOE\_peptide

FT

EP-743364-A2.

PD 20-NOV-1996.

PF 17-MAY-1996; 401084.

PR 18-MAY-1995; FR-005914.

PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.

PI Narwa R, Roques P;

DR WPI: 96-507733/51.

DR P-PSDB: W06615.

PT Human immunodeficiency virus p17 gene fragments, derived proteins

PT and antibodies - useful for assessing the risk of maternal

PT transmission of HIV-1 infection  
PS Claim 3; Page 26; 46pp; French.

CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, CTA, GTG or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-foetal transmission of HIV-1, using

CC standard hybridisation or immuno assays. The presence of such sequences  
 CC in maternal blood is strongly correlated with transmission of infection.  
 SQ Sequence 54 BP; 30 A; 6 C; 13 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
 Best Local Similarity 74.1%; Pred. No. 2.1e+02;

Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 agatgccaaagcaatcaacgtaagaa 27  
 ||| | ||||| ||| |||||  
 Db 21 AGAGGAAGACCAACCAAGTAAGAA 47

## RESULT 15

T43657  
 ID T43657 standard; DNA; 54 BP.

AC T43657;

DT 19-AUG-1997 (first entry)  
 DE HIV-1 matrix protein p17 gene fragment 4541.

KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
 detection; maternal transmission; hybridisation assay; immunoassay; ss.

OS Human immunodeficiency virus type 1.  
 PN EP-743364-A2.

PD 20-NOV-1996.

PF 17-MAY-1996; 401084.

PR 18-MAY-1995; FR-005914.

PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.

PI Narwa R. Roques P;

DR MPI; 96-507733/51.

PT Human immunodeficiency virus p17 gene fragments, derived proteins  
 and antibodies - useful for assessing the risk of maternal

PT Transmission of HIV-1 infection

PS Claim 3; Page 27; 46pp; French.

CC This sequence is a specifically claimed example of 21-90 nucleotide  
 long nucleic acid fragments, derived from the gene encoding part of the

p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
 least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;

Y6 = ATA, TTA, CTG, GTA, CTA, GTG or ATG; Y7 = GAG or GAA; Y8 = GAA or  
 CAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or

CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
 = AAA or AAG, then Y6 is not ATR, Y7 is not GAG or GAA, Y8 is not GAA,

Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.

CC The new nucleic acid fragments can be used as reagents for determining  
 and assessing the risk of maternal-foetal transmission of HIV-1, using

CC standard hybridisation or immuno assays. The presence of such sequences  
 in maternal blood is strongly correlated with transmission of infection.

Sequence 54 BP; 30 A; 4 C; 15 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
 Best Local Similarity 74.1%; Pred. No. 2.1e+02;

Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 agatgccaaagcaatcaacgtaagaa 27  
 ||| | ||||| ||| |||||  
 Db 21 AGAGGAAGACCAACCAAGTAAGAA 47

Search completed: June 4, 2000, 16:24:08  
 Job time: 28904 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 13:53:09 ; Search time 4521.53 Seconds  
(without alignments)  
25.100 Million cell updates/sec

Title: US-09-164-714-9

Perfect score: 28

Sequence: 1 agatgccagcaatcaatcaacgtaagaac 28

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026511650 residues

Total number of hits satisfying chosen parameters: 119968

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: gb\_est20:\*  
53: em\_est21:\*  
54: em\_est12:\*  
55: em\_est13:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est13:\*  
60: gb\_est14:\*  
61: gb\_est15:\*  
62: gb\_est16:\*  
63: gb\_est17:\*  
64: gb\_est18:\*  
65: em\_est17:\*  
66: em\_est18:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est19:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est13:\*  
78: em\_est14:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result

\*  
Query



found through the I.M.A.G.E. Consortium/LNL at:  
www.bio.lnl.gov/bdrp/image/image.html

Insert Length: 648 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 63.

#### FEATURES

Location/Qualifiers

1..63  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1251180"  
/clone\_lib="NCI CGAP GCBI"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD-),  
provided by Dr. Louis M. Staudt (NCI). Dr. David Allman  
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
primed with a Not I - oligo(dT) primer  
[5'-TGTACCAATCTGAAGTGGAGCGCCCTCATTTTTTTTTTTTTTTT-  
3']. Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT733 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 26 a 10 c 18 g 9 t  
ORIGIN

Query Match 54.3%; Score 15.2; DB 38; Length 63;  
Best Local Similarity 71.4%; Pred. No. 7.3e+03;  
Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 agatgccaaagcaatcaacgctaagaac 28  
||||| | | | | | | | | | | | | | | | | | | | |  
Db 15 AGATGCAGAGAACTCAAGCTAAATAC 42

RESULT 3  
AA244996 72 bp mRNA EST 12-MAR-1997  
LOCUS  
DEFINITION mw29d07.11 Guaywoodford Beier mouse kidney day 0 Mus musculus cDNA  
Clone IMAGE:656461 5', mRNA sequence.  
ACCESSION AA244996  
VERSION  
KEYWORDS  
SOURCE EST.  
ORGANISM house mouse.  
Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 72)

REFERENCE  
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisler, S., Kucada, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMI Mouse EST Project

TITLE Unpublished (1996)  
JOURNAL On Dec 30, 1996 this sequence version replaced gi:1530953.  
COMMENT Contact: Marra M/Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mousest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:402309  
Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 57.

#### FEATURES

Location/Qualifiers

1..72  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:656461"  
/clone\_lib="Guaywoodford Beier mouse kidney day 0"  
/tissue\_type="kidney"  
/dev\_stage="newborn (day 0)"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: kidney; Vector: Bluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. Average insert size: 1.0 kb; Uni-ZAP XR Vector;  
-5' adaptor sequence: 5' GAATTGGCGCAG 3' -3' adaptor  
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' Library provided  
Lisa Guay-woodford."

BASE COUNT 33 a 11 c 14 g 13 t 1 others  
ORIGIN

Query Match 54.3%; Score 15.2; DB 30; Length 72;  
Best Local Similarity 71.4%; Pred. No. 7.4e+03;  
Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 agatgccaaagcaatcaacgctaagaac 28  
||||| | | | | | | | | | | | | | | | | | | | |  
Db 45 AGATGCAGAGCTAAATTAAGTAGAGAC 72

RESULT 4  
AA746399 75 bp mRNA EST 27-JAN-1998  
LOCUS  
DEFINITION nw62c05.s1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:1251176 3',  
mRNA sequence.  
ACCESSION AA746399  
VERSION  
KEYWORDS  
SOURCE EST.  
ORGANISM human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 75)

REFERENCE  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.L.Strausberg@nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,  
Ph.D., Gerald Marti, M.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone Distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www.bio.lnl.gov/bdrp/image/image.html

Insert Length: 363 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 63.

#### FEATURES

Location/Qualifiers

1..75  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1251176"  
/clone\_lib="NCI CGAP GCBI"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for

germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-GTGTACCAATCTGAAGTGGAGCGCGCCCTCATTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTV3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 29 a 14 c 20 g 12 t

ORIGIN

Query Match 54.3%; Score 15.2; DB 38; Length 75;  
Best Local Similarity 71.4%; Pred. No. 7.4e+03;  
Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agatgccagcaaatcaacggtgaagaac 28  
||||| || ||||| ||||| ||  
Db 15 AGATGCAGAGAACTCAAGCTAAATAC 42

RESULT 5  
AA746402 75 bp mRNA EST 27-JAN-1998  
LOCUS n6262c11.s1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:1251188 3',  
DEFINITION mRNA sequence.  
ACCESSION AA746402  
VERSION AA746402.1 GI:2786388  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 75)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,  
Ph.D., Gerald Marti, M.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.

CDNA library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Cloned through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/dbp/imagen/image.html

Insert Length: 664 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 63.  
Location/Qualifiers

## FEATURES

source

1. .75  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1251188"  
/clone\_lib="NCI\_CGAP\_GCB1"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="Vector: pTV3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD-),  
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
primed with a Not I - oligo(dT) primer  
[5'-GTGTACCAATCTGAAGTGGAGCGCGCCCTCATTTTTTTTTTTTTTT-  
3']. Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pTV3 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 29 a 14 c 20 g 12 t

ORIGIN

Query Match 54.3%; Score 15.2; DB 38; Length 75;  
Best Local Similarity 71.4%; Pred. No. 7.4e+03;  
Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agatgccagcaaatcaacggtgaagaac 28  
||||| || ||||| ||||| ||  
Db 15 AGATGCAGAGAACTCAAGCTAAATAC 42

RESULT 6  
AA681983 83 bp mRNA EST 05-DEC-1997  
LOCUS vu76c07.r1 StrataGene mouse skin (#937313) Mus musculus cDNA clone  
DEFINITION IMAGE:1197324 5', mRNA sequence.  
ACCESSION AA681983  
VERSION AA681983.1 GI:2664123  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 83)  
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Thelning, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

WashU-HMI Mouse EST Project  
Unpublished (1996)  
On Sep 12, 1996 this sequence version replaced gi:1392917.

COMMENT Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
NCI:644420

Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 76.  
Location/Qualifiers

## FEATURES

source

1. .83  
/organism="Mus musculus"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1197324"  
/clone\_lib="Stratagene mouse skin (#937313)"  
/sex="females"  
/tissue\_type="whole skin"  
/dev\_stage="11 weeks old"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: skin; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT. Whole skin from 11 week old C57BL/6 female mice.  
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'  
adaptor sequence: 5' GATTTCGGCAGCAG 3' -3' adaptor  
sequence: 5' CTCGAGCTTTTTTTTTTTTTTTT 3"

## BASE COUNT

38 a 12 c 14 g 19 t

## ORIGIN

Query Match 54.3%; Score 15.2; DB 37; Length 83;  
Best Local Similarity 71.4%; Pred. No. 7.5e+03;

Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 agatgccagaatcaacgtaagaac 28  
 ||||| | ||||| | ||||| |  
 Db 22 AGATGAAGACTAATTAAGTAAGAAC 49

RESULT 7  
 D11978 89 bp mRNA EST 02-DEC-1992  
 LOCUS HUM0512E03 Liver Hepg2 cell line. Homo sapiens cDNA clone s12e03,  
 DEFINITION mRNA sequence.  
 D11978  
 ACCESSION D11978.1 GI:2148809  
 VERSION D11978.1 GI:2148809  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
 1 (bases 1 to 89)  
 Okubo,K., Horii,N., Matoba,R., Niyama,T., Fukushima,A., Kojima,Y.  
 and Matsubara,K.  
 Large scale cDNA sequencing for analysis of quantitative and  
 qualitative aspects of gene expression  
 Nature Genet. 2, 173-179 (1992)

JOURNAL  
 MEDLINE  
 COMMENT Contact: Kousaku Okubo, Naohiro Horii, Ryo Matoba, Toshiyuki  
 Niyama, Atsushi Fukushima, Yoko Kojima & Kenichi Matsubara  
 Institute for Molecular and Cellular Biology  
 Osaka University  
 1-3 Yamada-oka, Suita, Osaka 565, Japan.  
 Location/Qualifiers

FEATURES  
 source  
 1..89  
 /organism="Homo sapiens"  
 /db\_xref="GDB:D058180E"  
 /db\_xref="taxon:9606"  
 /clone="s12e03"  
 /clone\_lib="Liver Hepg2 cell line."  
 /lab\_host="E.coli"  
 /note="3'-directed regional cDNA library. Cleaved by MboI  
 and transformed into E.coli."

BASE COUNT 14 a 21 c 22 g 32 t

ORIGIN

Query Match 54.3%; Score 15.2; DB 20; Length 89;  
 Best Local Similarity 71.4%; Pred. No. 7.5e+03;  
 Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 agatgccagaatcaacgtaagaac 28  
 ||||| | ||||| | ||||| |  
 Db 73 AAATGCCAGACCAATCAACAGGAGGAC 46

RESULT 8  
 D12000/c 94 bp mRNA EST 02-DEC-1992  
 LOCUS HUM0513A08 Liver Hepg2 cell line. Homo sapiens cDNA clone s13a08,  
 DEFINITION mRNA sequence.  
 D12000  
 ACCESSION D12000.1 GI:2148823  
 VERSION D12000.1 GI:2148823  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
 1 (bases 1 to 94)  
 Okubo,K., Horii,N., Matoba,R., Niyama,T., Fukushima,A., Kojima,Y.  
 and Matsubara,K.  
 Large scale cDNA sequencing for analysis of quantitative and  
 qualitative aspects of gene expression  
 Nature Genet. 2, 173-179 (1992)

JOURNAL  
 MEDLINE

COMMENT Contact: Kousaku Okubo, Naohiro Horii, Ryo Matoba, Toshiyuki  
 Niyama, Atsushi Fukushima, Yoko Kojima & Kenichi Matsubara  
 Institute for Molecular and Cellular Biology  
 Osaka University  
 1-3 Yamada-oka, Suita, Osaka 565, Japan.  
 Location/Qualifiers

FEATURES  
 source  
 1..94  
 /organism="Homo sapiens"  
 /db\_xref="GDB:D058194E"  
 /db\_xref="taxon:9606"  
 /clone="s13a08"  
 /clone\_lib="Liver Hepg2 cell line."  
 /lab\_host="E.coli"  
 /note="3'-directed regional cDNA library. Cleaved by MboI  
 and transformed into E.coli."

BASE COUNT 18 a 21 c 21 g 33 t

ORIGIN

Query Match 54.3%; Score 15.2; DB 20; Length 94;  
 Best Local Similarity 71.4%; Pred. No. 7.5e+03;  
 Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 agatgccagaatcaacgtaagaac 28  
 ||||| | ||||| | ||||| |  
 Db 73 AAATGCCAGACCAATCAACAGGAGGAC 46

RESULT 9  
 AF088162/c 84 bp DNA GSS 10-DEC-1998  
 LOCUS AF088162 Homo sapiens chromosome 11 clone 49D4E3 map 11q13, genomic survey  
 DEFINITION sequence.  
 AF088162  
 ACCESSION AF088162.1 GI:3982872  
 VERSION AF088162.1 GI:3982872  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
 1 (bases 1 to 84)  
 Gerhardt,D.S. and Zhu,S.  
 A transcript map of an 800 kb region on human chromosome 11q13,  
 part of the candidate region for SCAs and BBS1  
 Hum. Genet. (1999) In press

JOURNAL  
 MEDLINE  
 COMMENT Direct Submission  
 Submitted (31-AUG-1998) Department of Genetics, Washington  
 University School of Medicine, 4566 Scott Avenue, Box 8232, St.  
 Louis, MO 63110, USA  
 Location/Qualifiers

FEATURES  
 source  
 1..84  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="11q13"  
 /clone="49D4E3"  
 /chromosome="11"

BASE COUNT 24 a 17 c 10 g 33 t

ORIGIN

Query Match 53.6%; Score 15; DB 83; Length 84;  
 Best Local Similarity 78.3%; Pred. No. 9.1e+03;  
 Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 5 gccagaatcaacgtaagaac 27  
 ||||| | ||||| | ||||| |  
 Db 34 GCCAAGCACATCAAAAGAACGAA 12

RESULT 10  
 AA930201

LOCUS	AA930201	81 bp	mRNA	EST	23-APR-1998
DEFINITION	vz12e03.t1 Strataegae mouse skin (#937313) Mus musculus cDNA clone IMAGE:1151740 5' similar to gb:U15647_c0s1 Mus musculus (MOUSE); , mRNA sequence.				
ACCESSION	AA930201				
VERSION	AA930201.1	GI:3079794			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Districh,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,T., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.				
TITLE	The Mashu-HMI Mouse EST Project				
JOURNAL	Unpublished (1996)				
COMMENT	On Jan 19, 1998 this sequence version replaced gi:2286367. Contact: Marra M/Mouse EST Project Mashu-HMI Mouse EST Project Washington University School of MedicineP 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LML ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:624948 Seq primer: -28ml3 rev1 EF from Amersham.				
FEATURES	Location/Qualifiers				
source	1. . 81				
	/organism="Mus musculus"				
	/strain="C57BL/6"				
	/db_xref="taxon:10090"				
	/clone="IMAGE:1151740"				
	/clone_lib="Strataegae mouse skin (#937313)"				
	/sex="females"				
	/tissue_type="whole skin"				
	/dev_stage="11 weeks old"				
	/lab_host="SOLR (kanamycin resistant)"				
	/note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: 0150 dt. Whole skin from 11 week old C57BL/6 female mice. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAAATCGCGACGAG 3' -3' adaptor sequence: 5' CTCGAGTGTGTGTGTGTGTGTGT 3'"				
BASE COUNT	35 a	21 c	9 g	16 t	
ORIGIN					
Query Match	52.9%	Score 14.8;	DB 40;	Length 81;	
Best Local Similarity	73.1%	Pred. No. 1.1e+04;			
Matches 19;	Conservative 0;	Mismatches 7;	Indels 0;	Gaps 0;	
QY	1	agatgcgaagcaatcaacgcytaa	26		
Db	41	AAATGCAATCAACAACACCTGACA	66		
RESULT 11					
A1536907					
LOCUS	A1536907	90 bp	mRNA	EST	12-MAY-1999
DEFINITION	tcl4a09.x1 NC1_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2179000 3' , mRNA sequence.				
ACCESSION	A1536907				
VERSION	A1536907.1	GI:4451042			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
	Eutheria; Primates; Catarrhini; Homiidae; Homo.				

REFERENCE	1 (bases 1 to 90)
AUTHORS	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	On May 18, 1998 this sequence version replaced g1:3138691. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert_Strausberg@nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <a href="http://www-bio.llnl.gov/bdrp/image/image.html">www-bio.llnl.gov/bdrp/image/image.html</a>
FEATURES	
SOURCE	Insert Length: 1143 Std Error: 0.00 Seq primer: -40UP from Glbco High quality sequence stop: 85 POLYA-No. Location/Qualifiers 1..90 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2179000" /clone_1id="NCI-CGAP-UT2" /tissue_type="moderately-differentiated endometrial adenocarcinoma, 3 pooled tumors" /lab_host="DH10B" /note="Organ: uterus; Vector: PCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.85 kb. Life Technologies catalog #: 11539-012"
BASE COUNT	34 a 16 c 13 g 27 t
ORIGIN	
Query Match	52.1%; Score 14.6; DB 47; Length 90;
Best Local Similarity	81.0%; Pred. No. 1.4e+04;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
Ox	7 caagcaaatcaacgtaagaa 27                   Db 17 CAAGCACAACAACGTAAGAA 37
RESULT 12	
AA161708	60 bp mRNA EST 17-DEC-1996
LOCUS	MAAFCE9PF053 Brugia malayi adult female cDNA (S9696MLW-BMAF) Brugia
DEFINITION	malayi cDNA clone AFCE9F05 5', mRNA sequence.
ACCESSION	AA161708
VERSION	AA161708.1 GI:1737540
KEYWORDS	EST.
SOURCE	Brugia malayi.
ORGANISM	Brugia malayi Eukaryota; Metazoa; Nematoda; Secernentea; Spirurina; Spirurida; Filarioida; Onchocercidae; Brugia. 1 (bases 1 to 60) Blaxter,M.L., Waterfall,M., Daub,J.J., Lizotte,M., Baron,L. and Jones,S.J. Genes expressed in adult female Brugia malayi Unpublished (1996) On May 18, 1995 this sequence version replaced g1:811474. Contact: Blaxter ML Institute of Cell, Animal and Population Biology University of Edinburgh Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JF, UK. Tel.: +44 131 650 6760 Fax: +44 131 670 5450
TITLE	
JOURNAL	
COMMENT	

Email: mark.blaxter@ed.ac.uk  
The ABI trace of this sequence can be viewed at  
<http://www.sanger.ac.uk/brugia/ABC/MBACF3P50573.html>  
Seq primer: T3.

## FEATURES

Location/Qualifiers  
1..60

/organism="Brugia malayi"  
/db\_xref="taxon:6279"  
/clone="AFCE3P05"  
/clone\_1lb="Brugia malayi adult female cDNA  
(SAM96MLW-BMAF)"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: Lambda uni-ZAP XR; Site\_1: Eco RI; Site\_2:  
Xho I; Lymphatic filarial nematode parasite of humans.  
mRNA was prepared from approximately 50 adult females  
isolated from the peritoneal cavity of jirds and  
converted to double-stranded cDNA using reverse  
transcriptase and oligo(dT) followed by RNase H and DNA  
pol I. The library has 5 x 10<sup>6</sup> independent recombinants  
and the average insert size is ~900bp. The library was  
constructed by Michelle Lizotte-Waniewski. The  
library is available from Dr.S.A.Williams, email:  
genomesmith.edu"

BASE COUNT 10 a 4 c 8 g 38 t  
ORIGIN

Query Match 51.4%; Score 14.4; DB 29; Length 60;  
Best Local Similarity 75.0%; Pred. No. 1.6e+04;  
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 atgccaaagcaaatcaacgtaaga 26  
|| ||||| ||| |||  
Db 25 ATACCAAGCAAAACAAATTAAGA 2

RESULT 13  
AM057193 82 bp mRNA EST 29-SEP-1999  
LOCUS ca02a05.y1 C elegans fem3 Q23 S1 Caenorhabditis elegans cDNA 5'  
DEFINITION similar to gb:R13H9.2 (ELEGANS); WP:R09C12.7 CE02603 ;, mRNA  
sequence.  
ACCESSION AM057193.1 GI:5932832  
VERSION AM057193  
KEYWORDS EST.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans.

REFERENCE  
AUTHORS Ward,S., Smith,H., Clifton,S., Marra,M., Hillier,L., Kucaba,T.,  
1 (bases 1 to 82)  
Rhabdilitia; Rhabdilitidae; Rhabdilitidae; Peloderinae; Caenorhabditis.

Ward,S., Smith,H., Clifton,S., Marra,M., Hillier,L., Kucaba,T.,  
Bowers,Y., Person,B., Swaller,T., Steptoe,M., Gibbons,M.,  
Harvey,N., Ritter,E., Jackson,Y., McCann,R., Waterston,R. and  
Wilson,R.

TITLE Dofarizona-Washu C. elegans EST project  
JOURNAL Unpublished (1999)  
COMMENT On Dec 20, 1995 this sequence version replaced gi:1135389.

CONTACT: Samuel Ward, Ph.D.  
Dofarizona-Washu C. elegans EST project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: est@wustl.edu  
Contact Harold Smith (hes@u.arizona.edu) for further information  
relating to organism, libraries, or clone availability.  
Seq primer: -40UP from GIBCO.

FEATURES  
SOURCE 1..82  
Location/Qualifiers

/organism="Caenorhabditis elegans"

/db\_xref="taxon:6239"  
/clone\_1lb="C. elegans fem3 Q23 S1"  
/lab\_host="DH5alpha cells"  
/note="Vector: pBluescript II SK+; Site\_1: XhoI; Site\_2:  
NotI; This C elegans library was made from fem-3(q23)  
worms (produce only sperm at 25 C). cDNA was generated  
via oligo (dT) priming and directionally cloned into  
pBluescript II SK+ vector with a modified polylinker.  
This library was subtracted with a second fem-1(hc17)  
(produce only oocytes at 25 C) library to enrich this  
library for sperm specific genes. First strand cDNA  
synthesis was primed with a NotI-15T oligo (sequence  
5'-CAGTACTGCTGATCGCGCGCCCTTTTCTTTTCTTTT-3'). After  
2nd strand synthesis, XhoI-EcoRI adaptors (Stratagene  
#901120) were ligated, and cDNA was digested with NotI and  
ligated into XhoI-NotI digested vector."

BASE COUNT 20 a 22 c 22 g 18 t  
ORIGIN

Query Match 51.4%; Score 14.4; DB 64; Length 82;  
Best Local Similarity 75.0%; Pred. No. 1.6e+04;  
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 agatgccaaagcaaatcaacgtaaa 24  
|| ||||| ||| |||  
Db 9 AGCTGCCAAGCAATTCGCCCGTTA 32

RESULT 14  
AA916294 87 bp mRNA EST 17-JUN-1998  
LOCUS on22b10.s1 NCI-CGAP.Lu5 Homo sapiens cDNA clone IMAGE:1557403 3'  
DEFINITION similar to gb:X61123\_rnal BTG1 PROTEIN (HUMAN);, mRNA sequence.  
ACCESSION AA916294  
VERSION AA916294.1 GI:3055686  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eukaryota; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 87)  
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On May 5, 1995 this sequence version replaced gi:797728.

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILM at:  
[www-bio.llnl.gov/dbp/image/image.html](http://www-bio.llnl.gov/dbp/image/image.html)

Trace considered overall poor quality  
Insert Length: 990 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 1.

FEATURES  
SOURCE 1..87  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1557403"  
/clone\_1lb="NCI-CGAP.Lu5"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: p7T3D-Pac (Pharmacia) with a  
modified polylinker; 1st strand cDNA was prepared from





---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:09:41 ; Search time 244.64 Seconds

(Without alignments)  
14.877 Million cell updates/sec

Title: US-09-164-714-9

Perfect score: 28

Sequence: 1 agatgccaaagcaatcaacggtagaac 28

## Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 6492525 residues

Total number of hits satisfying chosen parameters: 375880

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/6.PCTUS.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	57.1	79	2	US-08-446-908-8
2	16	57.1	79	2	US-08-231-205A-8
3	16	57.1	79	4	US-08-871-161-8
4	15.8	56.4	54	4	US-08-649-991-10
5	15.8	56.4	54	4	US-08-649-991-11
6	15.8	56.4	54	4	US-08-649-991-12
7	15.8	56.4	54	4	US-08-649-991-13
8	15.8	56.4	54	4	US-08-649-991-14
9	15.8	56.4	54	4	US-08-649-991-15
10	15.8	56.4	54	4	US-08-649-991-16
11	15.8	56.4	54	4	US-08-649-991-17
12	15.8	56.4	54	4	US-08-649-991-18
13	15.8	56.4	54	4	US-08-649-991-20
14	15.8	56.4	54	4	US-08-649-991-22
15	15.8	56.4	54	4	US-08-649-991-23
16	15.8	56.4	54	4	US-08-649-991-25
17	15.8	56.4	54	4	US-08-649-991-26
18	15.8	56.4	54	4	US-08-649-991-27
19	15.8	56.4	54	4	US-08-649-991-119
20	15.8	56.4	54	4	US-08-649-991-120
21	15.8	56.4	54	4	US-08-649-991-121
22	15.8	56.4	54	4	US-08-649-991-122
23	15.8	56.4	54	4	US-08-649-991-123
24	15.8	56.4	54	4	US-08-649-991-124
25	15.8	56.4	54	4	US-08-649-991-125
26	15.8	56.4	54	4	US-08-649-991-126
27	15.8	56.4	54	4	US-08-649-991-128

28	15.8	56.4	54	4	US-08-649-991-129	Sequence 129, App
29	15	53.6	31	1	US-08-390-850-43	Sequence 43, Appl
30	15	53.6	31	2	US-08-435-634-43	Sequence 43, Appl
31	14.8	52.9	53	3	US-08-811-492-120	Sequence 120, App
32	14.8	52.9	54	4	US-08-649-991-21	Sequence 21, Appl
33	14.8	52.9	54	4	US-08-649-991-24	Sequence 24, Appl
34	14.8	52.9	54	4	US-08-649-991-31	Sequence 31, Appl
35	14.8	52.9	54	6	PCT-US94-09700-35	Sequence 35, Appl
36	14.8	52.9	65	2	US-08-273-594-29	Sequence 29, Appl
37	14.6	52.1	69	3	US-07-916-098A-26	Sequence 26, Appl
38	14.6	52.1	69	3	US-07-916-098A-31	Sequence 31, Appl
39	14.2	50.7	50	4	US-08-850-049-5	Sequence 5, Appl1
40	14.2	50.7	50	4	US-08-050-478-5	Sequence 5, Appl1
41	14.2	50.7	54	4	US-08-649-991-9	Sequence 9, Appl1
42	14.2	50.7	54	4	US-08-649-991-19	Sequence 19, Appl
43	14.2	50.7	57	7	5514566-17	Patent No. 5514566
44	14.2	50.7	99	2	US-08-463-115-69	Sequence 69, Appl
45	14.2	50.7	99	2	US-08-465-388-69	Sequence 69, Appl

## ALIGNMENTS

RESULT 1  
US-08-446-908-8  
Sequence 8, Application US/08446908  
Patent No. 5705149  
GENERAL INFORMATION:  
APPLICANT: Namen, Anthony E.  
APPLICANT: Goodwin, Raymond G.  
APPLICANT: Lupton, Stephen D.  
APPLICANT: Mochizuki, Diane Y.  
TITLE OF INVENTION: Interleukin-7 and Antibodies Reactive  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: WA  
COUNTRY: US  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple 7.1  
SOFTWARE: Microsoft Word, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446,908  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/231,205  
FILING DATE: 21-APR-1994  
APPLICATION NUMBER: US 07/957,649  
FILING DATE: 06-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/511,438  
FILING DATE: 13-APR-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/255,209  
FILING DATE: 07-OCT-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/113,566  
FILING DATE: 26-OCT-1987  
ATTORNEY/AGENT INFORMATION:  
NAME: Seese, Kathryn A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2104-D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
INFORMATION FOR SEQ ID NO: 8:

```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 79 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 19..60
; US-08-446-908-8

Query Match          57.1%; Score 16; DB 2; Length 79;
Best Local Similarity 79.2%; Pred. No. 50;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy      2  gatgccagaacaatcacggtgaag 25
Db      34  GATGACAAGGAATCGAAGGTAGG 57

RESULT 2
; Sequence 8, Application US/08231205A
; Patent No. 571585
; GENERAL INFORMATION:
; APPLICANT: Namen, Anthony E.
; APPLICANT: Goodwin, Raymond G.
; APPLICANT: Lupton, Stephen D.
; APPLICANT: Mochizuki, Diane Y.
; TITLE OF INVENTION: Interleukin-7 and Antibodies Reactive
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: US
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: Apple Macintosh
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,205A
; FILING DATE: 21-APR-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/957,649
; FILING DATE: 06-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,438
; FILING DATE: 13-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/255,209
; FILING DATE: 07-OCT-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/113,566
; FILING DATE: 26-OCT-1987
; ATTORNEY/AGENT INFORMATION:
; NAME: Seese, Kathryn A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2104-D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 79 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 19..60
; US-08-231-205A-8

Query Match          57.1%; Score 16; DB 2; Length 79;
Best Local Similarity 79.2%; Pred. No. 50;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy      2  gatgccagaacaatcacggtgaag 25
Db      34  GATGACAAGGAATCGAAGGTAGG 57

RESULT 3
; Sequence 8, Application US/08871161
; Patent No. 5965122
; GENERAL INFORMATION:
; APPLICANT: Namen, Anthony E.
; APPLICANT: Goodwin, Raymond G.
; APPLICANT: Lupton, Stephen D.
; APPLICANT: Mochizuki, Diane Y.
; TITLE OF INVENTION: Interleukin-7 and Antibodies Reactive
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: US
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: Apple Macintosh
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871,161
; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/446,908
; FILING DATE: 22-MAY-1995
; APPLICATION NUMBER: US 08/231,205
; FILING DATE: 21-APR-1994
; APPLICATION NUMBER: US 07/957,649
; FILING DATE: 06-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,438
; FILING DATE: 13-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/255,209
; FILING DATE: 07-OCT-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/113,566
; FILING DATE: 26-OCT-1987
; ATTORNEY/AGENT INFORMATION:
; NAME: Seese, Kathryn A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2104-D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 79 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

```

;      TOPOLOGY: linear
;      MOLECULE TYPE: CDNA
;      FEATURE:
;      NAME/KEY: CDS
;      LOCATION: 19..60
;
US-08-871-161-8

```

ery Match	57.1%;	Score 16;	DB 4;	Length 79;
st Local Similarity	79.2%;	Pred. No. 50;		
tches 19; Conservative	0;	Mismatches	5;	Indels

```
Qy      2 gatgccaaacaatcaacgtaag 25
          |||| |||| |||| | |||| |
Db     34 GATGACAAGGAATCGAAGGTAGC 57
```

RESULT 4  
US-08-649-991-10  
; Sequence 10, Application US/08649991  
; Patent No 5010463

TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE  
 TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR  
 TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF  
 TITLE OF INVENTION: MATERNOTOERIAL TRANSMISSION OF HIV-1  
 NUMBER OF SEQUENCES: 130  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP  
 STREET: 1800 M Street, N.W.  
 CITY: Washington

```

; ZIP: 20036-5869
; COMPUTER READABLE FORM:
;

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; ***** ADDITIONAL DATA *****

```

; APPLICATION NUMBER: US/08/649,991  
 ; FILING DATE: 17-MAY-1996  
 ; CLASSIFICATION: 435  
 ; PRIORITY INFORMATION DATE:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9505914  
FILING DATE: 18-MAY-1995  
ATTORNEY/AGENT INFORMATION:

NAME: Adler Reid G.  
REGISTRATION NUMBER: 30,988  
REFERENCE/DOCKET NUMBER: ORES-5003  
TELECOMMUNICATION INFORMATION:  
300 475 5000

TELEPHONE: 202-467-7000  
TELEFAX: 202-467-7176  
INFORMATION FOR SEQ ID NO: 10;  
SEQUENCE CHARACTERISTICS:

LENGTH: 54 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-649-991-10

Query Match	Score	DB	Length
Best Local Similarity	74.1%;	Pred.	0.57;
Matches	20;	Conservative	0;
		Mismatches	7;
		Indels	

Db 21 AGAGGAGAGCAAAACAAAGTAAGAA 47

RESULT 5  
US-08-649-991-11  
; Sequence 11, Application US/08649991  
; Patent No. 5919462  
; GENERAL INFORMATION:

1	APPLICANT:	Rogues, Plerie
2	TITLE OF INVENTION:	NUCLEIC ACID FRAGMENTS DERIVED FROM THE
3	TITLE OF INVENTION:	HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR
4	TITLE OF INVENTION:	APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF
5	TITLE OF INVENTION:	MATERNOFETAL TRANSMISSION OF HIV-1

ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP  
STREET: 1800 M Street, N.W.

COUNTRY: USA  
ZIP: 20036-5869  
COUNTRY: BRAZIL

```

; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```
; SOFTWARE: Patent Release #1.0,  
;  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/649,991  
; FILING DATE: 17-MAY-1996  
; CREATING USER: ACF
```

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9505914  
FILING DATE: 18-MAY-1995  
RECORDING/CLASSIFICATION INFORMATION:

ATTORNEY/AGENT INFORMATION:  
NAME: Adler, Reid G.  
REGISTRATION NUMBER: 30,988  
REFERENCE/DOCKET NUMBER: ORES-5003

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-467-7000  
TELEFAX: 202-467-7176  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:

SEQUENCE CHARACTERISTICS:  
LENGTH: 54 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
MODIFICATIONS: none

TOPOLOGY: Linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-649-991-11

Query Match	56.4%	Score 15.8	DB 4	Length 54
Best Local Similarity	74.1%	Pred. No. 57		
Matches 20; Conservative	0	Mismatches 7	Indels 0	Gaps 0

QY 1 agatgcgaagcaaatcaacggttaagaa 27  
||| | ||||| ||| ||||| |||  
Db 21 AGAGGAAGAGCAAAAACAAAAGTAAGAA 47

RESULT 6  
US-08-649-991-12  
: Sequence 12, Application US/08649991

```

; Patent No. 5919462
;
; GENERAL INFORMATION:
;
; APPLICANT: Narwa, Remy
;
; APPLICANT: Roques, Pierre
;
;

```

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP  
;; STREET: 1800 M Street, N.W.

```

; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5869
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,991
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9505914
; FILING DATE: 18-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler, Reid G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: ORES-5003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-467-7000
; TELEFAX: 202-467-7176
; INFORMATION FOR SEQ. ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-649-991-12

```

```

Query Match          56.4%; Score 15.8; DB 4; Length 54;
Best Local Similarity 74.1%; Pred. No. 57;
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

QY 1 agatgccaaagcaatcaacgctaagaa 27
    ||| | ||||| ||| |||||
Db 21 AGAGGAAGACCAAAACAAAGTAAGAA 47

```

```

RESULT 7
US-08-649-991-13
; Sequence 13, Application US/08649991
; Patent No. 5919462
; GENERAL INFORMATION:
; APPLICANT: Narwa, Remy
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE
; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR
; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF
; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP
; STREET: 1800 M Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5869
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,991
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9505914
; FILING DATE: 18-MAY-1995

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Adler, Reid G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: ORES-5003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-467-7000
; TELEFAX: 202-467-7176
; INFORMATION FOR SEQ. ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-649-991-13

```

```

Query Match          56.4%; Score 15.8; DB 4; Length 54;
Best Local Similarity 74.1%; Pred. No. 57;
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

QY 1 agatgccaaagcaatcaacgctaagaa 27
    ||| | ||||| ||| |||||
Db 21 AGAGGAAGACCAAAACAAAGTAAGAA 47

```

```

RESULT 8
US-08-649-991-14
; Sequence 14, Application US/08649991
; Patent No. 5919462
; GENERAL INFORMATION:
; APPLICANT: Narwa, Remy
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE
; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR
; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF
; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP
; STREET: 1800 M Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5869
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,991
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9505914
; FILING DATE: 18-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler, Reid G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: ORES-5003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-467-7000
; TELEFAX: 202-467-7176
; INFORMATION FOR SEQ. ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-649-991-14

```

Query Match 56.4%; Score 15.8; DB 4; Length 54;  
Best Local Similarity 74.1%; Pred. No. 57;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agatgccaaagcaatcaacgtaagaa 27  
||| | ||||| ||| |||||  
Db 21 AGAGGAAGAGCAAAACAAAGTAGAA 47

## RESULT 9

US-08-649-991-15  
; Sequence 15, Application US/08649991  
; Patent No. 5919462  
; GENERAL INFORMATION:  
; APPLICANT: Narwa, Remy  
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE  
; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR  
; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF  
; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1  
; NUMBER OF SEQUENCES: 130  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP  
; STREET: 1800 M Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036-5869  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/649,991  
; FILING DATE: 17-MAY-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 9505914  
; FILING DATE: 18-MAY-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Adler, Reid G.  
; REGISTRATION NUMBER: 30,988  
; REFERENCE/DOCKET NUMBER: ORES-5003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-467-7000  
; TELEFAX: 202-467-7176  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 54 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-649-991-15

Query Match 56.4%; Score 15.8; DB 4; Length 54;  
Best Local Similarity 74.1%; Pred. No. 57;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agatgccaaagcaatcaacgtaagaa 27  
||| | ||||| ||| |||||  
Db 21 AGAGGAAGAGCAAAACAAAGTAGAA 47

RESULT 10  
US-08-649-991-16  
; Sequence 16, Application US/08649991  
; Patent No. 5919462  
; GENERAL INFORMATION:  
; APPLICANT: Narwa, Remy

APPLICANT: Roques, Piere  
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE  
; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR  
; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF  
; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1  
; NUMBER OF SEQUENCES: 130  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP  
; STREET: 1800 M Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036-5869  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/649,991  
; FILING DATE: 17-MAY-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 9505914  
; FILING DATE: 18-MAY-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Adler, Reid G.  
; REGISTRATION NUMBER: 30,988  
; REFERENCE/DOCKET NUMBER: ORES-5003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-467-7000  
; TELEFAX: 202-467-7176  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 54 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-649-991-16

Query Match 56.4%; Score 15.8; DB 4; Length 54;  
Best Local Similarity 74.1%; Pred. No. 57;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agatgccaaagcaatcaacgtaagaa 27  
||| | ||||| ||| |||||  
Db 21 AGAGGAAGAGCAAAACAAAGTAGAA 47

RESULT 11  
US-08-649-991-17  
; Sequence 17, Application US/08649991  
; Patent No. 5919462  
; GENERAL INFORMATION:  
; APPLICANT: Narwa, Remy  
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE  
; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR  
; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF  
; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1  
; NUMBER OF SEQUENCES: 130  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP  
; STREET: 1800 M Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036-5869  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

US-08-649-991-17  
; Sequence 17, Application US/08649991  
; Patent No. 5919462  
; GENERAL INFORMATION:  
; APPLICANT: Narwa, Remy  
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE  
; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR  
; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF  
; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1  
; NUMBER OF SEQUENCES: 130  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP  
; STREET: 1800 M Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036-5869  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,991
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9505914
; FILING DATE: 18-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler, Reid G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: ORES-5003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-467-7000
; TELEFAX: 202-467-7176
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-649-991-17

Query Match          56.4%; Score 15.8; DB 4; Length 54;
Best Local Similarity 74.1%; Pred. No. 57;
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agatgccaagcaaatcaacgtaagaa 27
    ||| | ||||| ||| ||||| |||
DB 21 AGAGGAAGAGCAAAACAAAGTAAGAA 47

RESULT 12
US-08-649-991-18
; Sequence 18, Application US/08649991
; Patent No. 5919462
; GENERAL INFORMATION:
; APPLICANT: Narwa, Remy
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE
; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR
; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF
; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP
; STREET: 1800 M Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5869
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,991
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9505914
; FILING DATE: 18-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler, Reid G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: ORES-5003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-467-7000
; TELEFAX: 202-467-7176
```

```

; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-649-991-18

Query Match          56.4%; Score 15.8; DB 4; Length 54;
Best Local Similarity 74.1%; Pred. No. 57;
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agatgccaagcaaatcaacgtaagaa 27
    ||| | ||||| ||| ||||| |||
DB 21 AGAGGAAGAGCAAAACAAAGTAAGAA 47

RESULT 13
US-08-649-991-20
; Sequence 20, Application US/08649991
; Patent No. 5919462
; GENERAL INFORMATION:
; APPLICANT: Narwa, Remy
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE
; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR
; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF
; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP
; STREET: 1800 M Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5869
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,991
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9505914
; FILING DATE: 18-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler, Reid G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: ORES-5003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-467-7000
; TELEFAX: 202-467-7176
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-649-991-20

Query Match          56.4%; Score 15.8; DB 4; Length 54;
Best Local Similarity 74.1%; Pred. No. 57;
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agatgccaagcaaatcaacgtaagaa 27
    ||| | ||||| ||| ||||| |||
```



Db 21 AGAGGAGAGCAAAACAAGTAGAA 47

## RESULT 14

US-08-649-991-22

; Sequence 22, Application US/08649991

; Patent No. 5919462

; GENERAL INFORMATION:

; APPLICANT: Narwa, Remy

; APPLICANT: Roques, Piere

; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE

; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR

; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF

; TITLE OF INVENTION: MATERNOCERIAL TRANSMISSION OF HIV-1

; NUMBER OF SEQUENCES: 130

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP

; STREET: 1800 M Street, N.W.

; City: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20036-5869

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/649,991

; FILING DATE: 17-MAY-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: FR 9505914

; FILING DATE: 18-MAY-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Adler, Reid G.

; REGISTRATION NUMBER: 30,988

; REFERENCE/DOCKET NUMBER: ORES-5003

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-467-7000

; TELEFAX: 202-467-7176

; INFORMATION FOR SEQ ID NO: 22:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 54 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; US-08-649-991-22

Query Match 56.4%; Score 15.8; DB 4; Length 54;

Best Local Similarity 74.1%; Pred. No. 57;

Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agatgccagcaaatcaacggttaagaa 27

Db 21 AGAGGAGAGCAAAACAAGTAGAA 47

## RESULT 15

US-08-649-991-23

; Sequence 23, Application US/08649991

; Patent No. 5919462

; GENERAL INFORMATION:

; APPLICANT: Narwa, Remy

; APPLICANT: Roques, Piere

; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE

; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR

; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF

; TITLE OF INVENTION: MATERNOCERIAL TRANSMISSION OF HIV-1

; NUMBER OF SEQUENCES: 130

; CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP

STREET: 1800 M Street, N.W.

City: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20036-5869

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/649,991

; FILING DATE: 17-MAY-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: FR 9505914

; FILING DATE: 18-MAY-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Adler, Reid G.

; REGISTRATION NUMBER: 30,988

; REFERENCE/DOCKET NUMBER: ORES-5003

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-467-7000

; TELEFAX: 202-467-7176

; INFORMATION FOR SEQ ID NO: 23:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 54 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; US-08-649-991-23

Query Match 56.4%; Score 15.8; DB 4; Length 54;

Best Local Similarity 74.1%; Pred. No. 57;

Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agatgccagcaaatcaacggttaagaa 27

Db 21 AGAGGAGAGCAAAACAAGTAGAA 47

Search completed: June 4, 2000, 16:09:42  
Job time: 28061 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:05:33 ; Search time 1236.38 Seconds

(without alignments)  
-22.031 Million cell updates/sec

Title: us-09-164-714-10

Perfect score: 28

Sequence: 1 gtcctacgcgtgattgcttgcgcacatc 28

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 356616

Minimum DB seq length: 0  
Maximum DB seq length: 100  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl : \*

1: gb\_bal : \*

2: gb\_bal : \*

3: gb\_bal : \*

4: gb\_bal : \*

5: gb\_bal : \*

6: gb\_bal : \*

7: gb\_bal : \*

8: gb\_bal : \*

9: gb\_bal : \*

10: gb\_bal : \*

11: gb\_bal : \*

12: gb\_bal : \*

13: gb\_bal : \*

14: gb\_bal : \*

15: gb\_bal : \*

16: gb\_bal : \*

17: gb\_bal : \*

18: gb\_bal : \*

19: gb\_bal : \*

20: gb\_bal : \*

21: gb\_bal : \*

22: gb\_bal : \*

23: gb\_bal : \*

24: gb\_bal : \*

25: gb\_bal : \*

26: gb\_bal : \*

27: gb\_bal : \*

28: gb\_bal : \*

29: gb\_bal : \*

30: gb\_bal : \*

31: gb\_bal : \*

32: gb\_bal : \*

33: gb\_bal : \*

34: gb\_bal : \*

35: gb\_bal : \*

36: gb\_bal : \*

37: gb\_bal : \*

38: gb\_bal : \*

39: gb\_bal : \*

40: gb\_bal : \*

41: gb\_bal : \*

42: gb\_bal : \*

43: gb\_bal : \*

44: gb\_bal : \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16	57.1	79	107076	107076 Sequence 2
2	16	57.1	79	183771	183771 Sequence 8
3	15.8	56.4	54	A57944	A57944 Sequence 10
4	15.8	56.4	54	A57945	A57945 Sequence 11
5	15.8	56.4	54	A57946	A57946 Sequence 12
6	15.8	56.4	54	A57947	A57947 Sequence 13
7	15.8	56.4	54	A57948	A57948 Sequence 14
8	15.8	56.4	54	A57949	A57949 Sequence 15
9	15.8	56.4	54	A57950	A57950 Sequence 16
10	15.8	56.4	54	A57951	A57951 Sequence 17
11	15.8	56.4	54	A57952	A57952 Sequence 18
12	15.8	56.4	54	A57953	A57953 Sequence 19
13	15.8	56.4	54	A57954	A57954 Sequence 20
14	15.8	56.4	54	A57955	A57955 Sequence 21
15	15.8	56.4	54	A57956	A57956 Sequence 22
16	15.8	56.4	54	A57957	A57957 Sequence 23
17	15.8	56.4	54	A57958	A57958 Sequence 24
18	15.8	56.4	54	A57959	A57959 Sequence 25
19	15.8	56.4	54	A57960	A57960 Sequence 26
20	15.8	56.4	54	A57961	A57961 Sequence 27
21	15.8	56.4	54	E05823	E05823 DNA encod1n
22	15.8	56.4	54	I37030	I37030 Sequence 43
23	15.8	56.4	54	I37030	I37030 Sequence 43
24	15.8	56.4	54	I37030	I37030 Sequence 43
25	15.8	56.4	54	I37030	I37030 Sequence 43
26	15.8	56.4	54	I37030	I37030 Sequence 43
27	15.8	56.4	54	I37030	I37030 Sequence 43
28	15.8	56.4	54	I37030	I37030 Sequence 43
29	15.8	56.4	54	I37030	I37030 Sequence 43
30	15.8	56.4	54	I37030	I37030 Sequence 43
31	15.8	56.4	54	I37030	I37030 Sequence 43
32	15.8	56.4	54	I37030	I37030 Sequence 43
33	15.8	56.4	54	I37030	I37030 Sequence 43
34	15.8	56.4	54	I37030	I37030 Sequence 43
35	15.8	56.4	54	I37030	I37030 Sequence 43
36	15.8	56.4	54	I37030	I37030 Sequence 43
37	15.8	56.4	54	I37030	I37030 Sequence 43
38	15.8	56.4	54	I37030	I37030 Sequence 43
39	15.8	56.4	54	I37030	I37030 Sequence 43
40	15.8	56.4	54	I37030	I37030 Sequence 43
41	15.8	56.4	54	I37030	I37030 Sequence 43
42	15.8	56.4	54	I37030	I37030 Sequence 43
43	15.8	56.4	54	I37030	I37030 Sequence 43
44	15.8	56.4	54	I37030	I37030 Sequence 43
45	15.8	56.4	54	I37030	I37030 Sequence 43

#### ALIGNMENTS

[illegible]

FEATURES	Source	Location/Qualifiers
COMMENT	1. .54	
FEATURES	location/Qualifiers	
BASE COUNT	31 a 4 c 13 g 6 t	
ORIGIN		
Query Match	56.4%; Score 15.8; DB 5; Length 54;	
Best Local Similarity	74.1%; Pred. No. 7.2e+03;	
Matches	20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;	
QY	2 ttctaccgttgattgcttgcatc 28	
DB	47 TTCCTACCTTGTGTTGCTGCTCCCT 21	
RESULT	5	
LOCUS	A57946/c 54 bp DNA	PAT 05-MAR-1998
DEFINITION	Sequence 12 from Patent EP0743364.	
ACCESSION	A57946	
VERSION	A57946.1 GI:3713716	
KEYWORDS	unidentified.	
SOURCE	unidentified.	
ORGANISM	unclassified.	
REFERENCE	1 (bases 1 to 54)	
AUTHORS	Narwa, R. and Roques, P.	
TITLE	Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of	
FEATURES	location/Qualifiers	
COMMENT	1. .54	
FEATURES	location/Qualifiers	
BASE COUNT	30 a 4 c 15 g 5 t	
ORIGIN		
Query Match	56.4%; Score 15.8; DB 5; Length 54;	
Best Local Similarity	74.1%; Pred. No. 7.2e+03;	
Matches	20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;	
QY	2 ttctaccgttgattgcttgcatc 28	
DB	47 TTCCTACCTTGTGTTGCTGCTCCCT 21	
RESULT	5	
LOCUS	A57946 54 bp DNA	PAT 05-MAR-1998
DEFINITION	Sequence 12 from Patent EP0743364.	
ACCESSION	A57946	
VERSION	A57946.1 GI:3713716	
KEYWORDS	unidentified.	
SOURCE	unidentified.	
ORGANISM	unclassified.	
REFERENCE	1 (bases 1 to 54)	
AUTHORS	Narwa, R. and Roques, P.	
TITLE	Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of	

JOURNAL HIV-1 mother-foetal transmission  
Patent: EP 0743364-A 12-20-NOV-1996;  
COMMISSARIAT ENERGIE ATOMIQUE (FR)  
Other publication FR 2734281 961122.  
FEATURES  
source 1..54  
/db\_xref="taxon:32644"  
Location/Qualifiers

BASE COUNT 31 a 4 c 14 g 5 t  
ORIGIN

Query Match 56.4%; Score 15.8; DB 5; Length 54;  
Best Local Similarity 74.1%; Pred. No. 7.2e+03;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ttctacgltgattgctgcatct 28  
||||| ||| ||||| |||  
Db 47 TTCTACTTTTGTTCCTCTCTCT 21

RESULT 6  
A57947/c 54 bp DNA PAT 05-MAR-1998  
LOCUS A57947 Sequence 13 from Patent EP0743364.  
DEFINITION A57947  
ACCESSION A57947  
VERSION A57947.1 GI:3713717  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 13-20-NOV-1996;  
COMMISSARIAT ENERGIE ATOMIQUE (FR)  
COMMENT Other publication FR 2734281 961122.  
FEATURES  
source 1..54  
/db\_xref="taxon:32644"  
Location/Qualifiers

BASE COUNT 31 a 4 c 13 g 6 t  
ORIGIN

Query Match 56.4%; Score 15.8; DB 5; Length 54;  
Best Local Similarity 74.1%; Pred. No. 7.2e+03;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ttctacgltgattgctgcatct 28  
||||| ||| ||||| |||  
Db 47 TTCTACTTTTGTTCCTCTCTCT 21

RESULT 7  
A57948/c 54 bp DNA PAT 05-MAR-1998  
LOCUS A57948 Sequence 14 from Patent EP0743364.  
DEFINITION A57948  
ACCESSION A57948  
VERSION A57948.1 GI:3713718  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 14-20-NOV-1996;  
COMMISSARIAT ENERGIE ATOMIQUE (FR)

COMMENT Other publication FR 2734281 961122.  
FEATURES  
source 1..54  
/db\_xref="taxon:32644"  
Location/Qualifiers

BASE COUNT 32 a 5 c 12 g 5 t  
ORIGIN

Query Match 56.4%; Score 15.8; DB 5; Length 54;  
Best Local Similarity 74.1%; Pred. No. 7.2e+03;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ttctacgltgattgctgcatct 28  
||||| ||| ||||| |||  
Db 47 TTCTACTTTTGTTCCTCTCTCT 21

RESULT 8  
A57949/c 54 bp DNA PAT 05-MAR-1998  
LOCUS A57949 Sequence 15 from Patent EP0743364.  
DEFINITION A57949  
ACCESSION A57949  
VERSION A57949.1 GI:3713719  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 15-20-NOV-1996;  
COMMISSARIAT ENERGIE ATOMIQUE (FR)  
COMMENT Other publication FR 2734281 961122.  
FEATURES  
source 1..54  
/db\_xref="taxon:32644"  
Location/Qualifiers

BASE COUNT 30 a 5 c 14 g 5 t  
ORIGIN

Query Match 56.4%; Score 15.8; DB 5; Length 54;  
Best Local Similarity 74.1%; Pred. No. 7.2e+03;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ttctacgltgattgctgcatct 28  
||||| ||| ||||| |||  
Db 47 TTCTACTTTTGTTCCTCTCTCT 21

RESULT 9  
A57950/c 54 bp DNA PAT 05-MAR-1998  
LOCUS A57950 Sequence 16 from Patent EP0743364.  
DEFINITION A57950  
ACCESSION A57950  
VERSION A57950.1 GI:3713720  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 16-20-NOV-1996;  
COMMISSARIAT ENERGIE ATOMIQUE (FR)  
COMMENT Other publication FR 2734281 961122.  
FEATURES  
source 1..54  
Location/Qualifiers

BASE COUNT 30 a 5 c 14 g 5 t  
ORIGIN

Query Match 56.4%; Score 15.8; DB 5; Length 54;  
Best Local Similarity 74.1%; Pred. No. 7.2e+03;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ttctacgctgattgctgcacatc 28  
||||| ||| ||| ||| |||  
Db 47 TTCTTACTTTGTGTGCTCTCTCT 21

RESULT 10  
A57951/c 54 bp DNA PAT 05-MAR-1998  
LOCUS A57951 Sequence 17 from Patent EP0743364.  
DEFINITION A57951  
ACCESSION A57951  
VERSION A57951.1 GI:3713721  
KEYWORDS  
SOURCE .  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 17 20-NOV-1996;  
COMMENT COMMISSARIAT ENERGIE ATOMIQUE (FR)  
FEATURES Other publication FR 2734281 961122.  
source Location/Qualifiers  
1. .54  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 31 a 4 c 14 g 5 t  
ORIGIN

Query Match 56.4%; Score 15.8; DB 5; Length 54;  
Best Local Similarity 74.1%; Pred. No. 7.2e+03;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ttctacgctgattgctgcacatc 28  
||||| ||| ||| ||| |||  
Db 47 TTCTTACTTTGTGTGCTCTCTCT 21

RESULT 11  
A57952/c 54 bp DNA PAT 05-MAR-1998  
LOCUS A57952 Sequence 18 from Patent EP0743364.  
DEFINITION A57952  
ACCESSION A57952  
VERSION A57952.1 GI:3713722  
KEYWORDS  
SOURCE .  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 18 20-NOV-1996;  
COMMENT COMMISSARIAT ENERGIE ATOMIQUE (FR)  
FEATURES Other publication FR 2734281 961122.  
source Location/Qualifiers  
1. .54  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 29 a 5 c 15 g 4 t 1 others

ORIGIN

Query Match 56.4%; Score 15.8; DB 5; Length 54;  
Best Local Similarity 74.1%; Pred. No. 7.2e+03;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ttctacgctgattgctgcacatc 28  
||||| ||| ||| ||| |||  
Db 47 TTCTTACTTTGTGTGCTCTCTCT 21

RESULT 12  
A57954/c 54 bp DNA PAT 05-MAR-1998  
LOCUS A57954 Sequence 20 from Patent EP0743364.  
DEFINITION A57954  
ACCESSION A57954  
VERSION A57954.1 GI:3713724  
KEYWORDS  
SOURCE .  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 20 20-NOV-1996;  
COMMENT COMMISSARIAT ENERGIE ATOMIQUE (FR)  
FEATURES Other publication FR 2734281 961122.  
source Location/Qualifiers  
1. .54  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 30 a 4 c 15 g 5 t  
ORIGIN

Query Match 56.4%; Score 15.8; DB 5; Length 54;  
Best Local Similarity 74.1%; Pred. No. 7.2e+03;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ttctacgctgattgctgcacatc 28  
||||| ||| ||| ||| |||  
Db 47 TTCTTACTTTGTGTGCTCTCTCT 21

RESULT 13  
A57956/c 54 bp DNA PAT 05-MAR-1998  
LOCUS A57956 Sequence 22 from Patent EP0743364.  
DEFINITION A57956  
ACCESSION A57956  
VERSION A57956.1 GI:3713726  
KEYWORDS  
SOURCE .  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 22 20-NOV-1996;  
COMMENT COMMISSARIAT ENERGIE ATOMIQUE (FR)  
FEATURES Other publication FR 2734281 961122.  
source Location/Qualifiers  
1. .54  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 30 a 6 c 14 g 4 t

Query Match 56.4%; Score 15.8; DB 5; Length 54;

Best Local Similarity 74.1%; Pred. No. 7.2e+03;

Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 ttcttaccgttgattgcttgcattc 28  
||||| ||| ||| |||

Db 47 TTCTTACTTTGTGTTGCTCTCTCT 21

## RESULT 14

A57957/c

LOCUS A57957 54 bp DNA

DEFINITION Sequence 23 from Patent EP0743364.

ACCESSION A57957

VERSION A57957.1 GI:3713727

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 54)

AUTHORS

Narwa,R. and Roques,P.

Nucleic acid fragments derived from the HIV-1 genome, corresponding

fragments and their application as reactives for risk evaluation of

HIV-1 mother-foetal transmission

Patent: EP 0743364-A 23 20-NOV-1996;

COMMISSARIAT ENERGIE ATOMIQUE (FR)

Other publication FR 2734281 961122.

COMMENT

FEATURES

source

1..54

/organism="unidentified"

/db\_xref="taxon:32644"

BASE COUNT

30 a 6 c 13 g 5 t

ORIGIN

Query Match 56.4%; Score 15.8; DB 5; Length 54;

Best Local Similarity 74.1%; Pred. No. 7.2e+03;

Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 ttcttaccgttgattgcttgcattc 28  
||||| ||| ||| |||

Db 47 TTCTTACTTTGTGTTGCTCTCTCT 21

## RESULT 15

A57959/c

LOCUS A57959 54 bp DNA

DEFINITION Sequence 25 from Patent EP0743364.

ACCESSION A57959

VERSION A57959.1 GI:3713729

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 54)

AUTHORS

Narwa,R. and Roques,P.

Nucleic acid fragments derived from the HIV-1 genome, corresponding

fragments and their application as reactives for risk evaluation of

HIV-1 mother-foetal transmission

Patent: EP 0743364-A 25 20-NOV-1996;

COMMISSARIAT ENERGIE ATOMIQUE (FR)

Other publication FR 2734281 961122.

COMMENT

FEATURES

source

1..54

/organism="unidentified"

/db\_xref="taxon:32644"

BASE COUNT

30 a 4 c 15 g 5 t

ORIGIN

Query Match 56.4%; Score 15.8; DB 5; Length 54;

Best Local Similarity 74.1%; Pred. No. 7.2e+03;

Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 ttcttaccgttgattgcttgcattc 28  
||||| ||| ||| |||

Db 47 TTCTTACTTTGTGTTGCTCTCTCT 21

Search completed: June 4, 2000, 16:05:33  
Job time: 27885 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:24:08 ; Search time 322.35 Seconds  
(without alignments)  
21.732 Million cell updates/sec

Title: US-09-164-714-10

Perfect score: 28  
Sequence: 1 gtcttaccgttgattgcttgcatct 28

Scoring table: IDENTITY\_NNC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 segs, 125096042 residues

Total number of hits satisfying chosen parameters: 431286

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	15.8	56.4	54	1 T43658	HIV-1 matrix prote
C 2	15.8	56.4	54	1 T43659	HIV-1 matrix prote
C 3	15.8	56.4	54	1 T43642	HIV-1 matrix prote
C 4	15.8	56.4	54	1 T43643	HIV-1 matrix prote
C 5	15.8	56.4	54	1 T43644	HIV-1 matrix prote
C 6	15.8	56.4	54	1 T43645	HIV-1 matrix prote
C 7	15.8	56.4	54	1 T43646	HIV-1 matrix prote
C 8	15.8	56.4	54	1 T43647	HIV-1 matrix prote
C 9	15.8	56.4	54	1 T43648	HIV-1 matrix prote
C 10	15.8	56.4	54	1 T43649	HIV-1 matrix prote
C 11	15.8	56.4	54	1 T43650	HIV-1 matrix prote
C 12	15.8	56.4	54	1 T43652	HIV-1 matrix prote
C 13	15.8	56.4	54	1 T43654	HIV-1 matrix prote
C 14	15.8	56.4	54	1 T43655	HIV-1 matrix prote
C 15	15.8	56.4	54	1 T43657	HIV-1 matrix prote
C 16	15.8	56.4	92	1 O53311	Hybrid antigen pro
C 17	15.2	54.3	38	1 N94509	Probe for N-term
C 18	15.2	54.3	38	1 V83523	PCR primer used to
C 19	15.2	54.3	38	1 V83525	PCR primer used to
C 20	15.2	54.3	97	1 T19416	Human gene signatu
C 21	15	53.6	31	1 O93504	Human stromelysin
C 22	14.8	52.9	47	1 X52594	Human genome biall
C 23	14.8	52.9	47	1 X52596	Human genome biall
C 24	14.8	52.9	47	1 X52546	Human genome biall
C 25	14.8	52.9	47	1 X52544	Human genome biall
C 26	14.8	52.9	53	1 V68220	PCR primer used to
C 27	14.8	52.9	54	1 O86781	SDI-1 mimetic prim
C 28	14.8	52.9	54	1 O74004	SDI-1 mimetic PCR
C 29	14.8	52.9	54	1 T43663	HIV-1 matrix prote
C 30	14.8	52.9	54	1 T43663	HIV-1 matrix prote
C 31	14.8	52.9	54	1 T43656	HIV-1 matrix prote
C 32	14.6	52.1	43	1 T01299	PCR primer Oligo D
C 33	14.6	52.1	69	1 O30897	Primer 312-69. New
C 34	14.6	52.1	69	1 O30892	Primer 312-64. New

35	14.6	52.1	100	1 T42735	Primer J089 for hu
36	14.4	51.4	60	1 T62581	PCR primer for cys
C 37	14.2	50.7	47	1 V68272	Pentacillin chryso
C 38	14.2	50.7	54	1 T43651	HIV-1 matrix prote
C 39	14.2	50.7	54	1 T43641	HIV-1 matrix prote
C 40	14.2	50.7	99	1 V10203	Stealth virus nucl
C 41	14.2	50.7	99	1 V12016	Stealth virus plas
C 42	14	50.0	99	1 T30883	Primer 13 for 95 k
C 43	13.8	49.3	23	1 O88336	PCR primer Ofg3042
C 44	13.8	49.3	30	1 O88341	PCR primer Ofg5988
C 45	13.8	49.3	37	1 V39939	Streptococcus pneu

## ALIGNMENTS

RESULT 1	T43658/c	T43658 standard; DNA; 54 BP.
ID	T43658:	
AC	T43658:	
DT	19-AUG-1997	(first entry)
DE	HIV-1 matrix protein p17 gene fragment 2754.	
KW	Human immunodeficiency virus; matrix protein p17; prognosis; probe;	
KW	detection; maternal transmission; hybridisation assay; immunoassay; ss.	
OS	Human immunodeficiency virus type 1.	
PN	EP-743364-A2.	
PD	20-NOV-1996.	
PF	17-MAY-1996; 401084.	
PR	18-MAY-1995; FR-005914.	
PA	(COMS ) COMMISSARIAT ENERGIE ATOMIQUE.	
PI	Narwa R. Roques P;	
DR	WPT: 96-507733/51.	
PT	Human immunodeficiency virus p17 gene fragments, derived proteins	
PT	and antibodies - useful for assessing the risk of maternal	
PT	transmission of HIV-1 infection	
PS	Claim 3; Page 27; 46pp; French.	
CC	This sequence is a specifically claimed example of 21-90 nucleotide	
CC	long nucleic acid fragments, derived from the gene encoding part of the	
CC	p17 HIV-1 matrix protein; the claimed nucleic acids must include at	
CC	least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;	
CC	Y6 = ATA, TTA, CTG, GTA, CTA, GTC or ATG; Y7 = GAG or GAA; Y8 = GAA or	
CC	AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or	
CC	CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5	
CC	= AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,	
CC	Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.	
CC	The new nucleic acid fragments can be used as reagents for determining	
CC	and assessing the risk of maternal-foetal transmission of HIV-1, using	
CC	standard hybridisation or immuno assays. The presence of such sequences	
CC	in maternal blood is strongly correlated with transmission of infection.	
CC	Sequence 54 BP; 30 A; 4 C; 15 G; 5 T;	
QY	2 ttcttaccgttgattgcttgcatct 28	
DB	47 ttcttaccgttgattgcttgcatct 21	
Query Match	56.4%; Score 15.8; DB 1; Length 54;	
Best Local Similarity	74.1%; Pred. No. 2.1e+02;	
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;		
RESULT 2	T43659/c	
ID	T43659 standard; DNA; 54 BP.	
AC	T43659:	
DT	19-AUG-1997	(first entry)
DE	HIV-1 matrix protein p17 gene fragment 2826.	
KW	Human immunodeficiency virus; matrix protein p17; prognosis; probe;	
KW	detection; maternal transmission; hybridisation assay; immunoassay; ss.	
OS	Human immunodeficiency virus type 1.	
PN	EP-743364-A2.	
PD	20-NOV-1996.	
PF	17-MAY-1996; 401084.	

PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R, Rogues P;  
DR WPI: 96-507733/51.  
PT Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal  
PT transmission of HIV-1 infection  
PS Claim 3; Page 27; 46pp; French.  
CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, GTA, CTA, GTG or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-fetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 30 A; 4 C; 15 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. NO. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 ttctaccgttgattgctgcacatc 28  
Db 47 TTCTTACTTTGTTTGCTCTTCCTCT 21

RESULT 3  
T43642/C  
ID T43642 standard; DNA: 54 BP.  
AC T43642.  
DT 19-AUG-1997 (first entry)  
DE HIV-1 matrix protein p17 gene fragment ARI.  
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
KW detection; maternal transmission; hybridisation assay; immunoassay; ss.  
CC Human immunodeficiency virus type 1.  
PN Ep-743364-A2.  
PD 20-NOV-1996.  
PF 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R, Rogues P;  
DR WPI: 96-507733/51.  
PT Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal  
PT transmission of HIV-1 infection  
PS Claim 3; Page 23; 46pp; French.  
CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, GTA, CTA, GTG or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-fetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 31 A; 4 C; 13 G; 6 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. NO. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 ttctaccgttgattgctgcacatc 28  
Db 47 TTCTTACTTTGTTTGCTCTTCCTCT 21

RESULT 4  
T43643/C  
ID T43643 standard; DNA: 54 BP.  
AC T43643.  
DT 19-AUG-1997 (first entry)  
DE HIV-1 matrix protein p17 gene fragment BOI.  
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
KW detection; maternal transmission; hybridisation assay; immunoassay; ss.  
FH Key location/Qualifiers  
FT mat\_peptide 1..54  
FT /\*tag= a  
FT /product= BOI\_peptide  
PN Ep-743364-A2.  
PD 20-NOV-1996.  
PF 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R, Rogues P;  
DR WPI: 96-507733/51.  
DR P-PSDB; W06610.  
PT Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal  
PT transmission of HIV-1 infection  
PS Claim 3; Page 23; 46pp; French.  
CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, GTA, CTA, GTG or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-fetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 30 A; 4 C; 15 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. NO. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 ttctaccgttgattgctgcacatc 28  
Db 47 TTCTTACTTTGTTTGCTCTTCCTCT 21

RESULT 5  
T43644/C  
ID T43644 standard; DNA: 54 BP.  
AC T43644.  
DT 19-AUG-1997 (first entry)  
DE HIV-1 matrix protein p17 gene fragment DUM.  
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
KW detection; maternal transmission; hybridisation assay; immunoassay; ss.  
PN Ep-743364-A2.  
PD 20-NOV-1996.  
PF 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R, Rogues P;  
DR WPI: 96-507733/51.  
PT Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal

PT transmission of HIV-1 infection  
PS Claim 3: Page 24; 46pp; French.  
CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, GTA, CTA, GTC or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAC, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-fetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 31 A; 4 C; 14 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 ttctaccgtgattgctgcacatc 28  
||||| ||| ||||| |||  
DB 47 TTCTTACTTTTGTGTTGCTCTCTCTCT 21

RESULT 6  
T43645/c  
ID T43645 standard; DNA: 54 BP.  
AC T43645;  
DE 19-AUG-1997 (first entry)  
DT HIV-1 matrix protein p17 gene fragment PAL.  
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
KW detection; maternal transmission; hybridisation assay; immunoassay; ss.  
OS Human immunodeficiency virus type 1.  
PN EP-743364-A2.  
PD 20-NOV-1996.  
PF 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R. Roques P;  
DR WPI: 96-507733/51.

PT Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal  
PT transmission of HIV-1 infection  
PS Claim 3: Page 24; 46pp; French.

CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, GTA, CTA, GTC or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAC, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-fetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 31 A; 4 C; 13 G; 6 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 ttctaccgtgattgctgcacatc 28  
||||| ||| ||||| |||  
DB 47 TTCTTACTTTTGTGTTGCTCTCTCTCT 21

RESULT 7

T43646/c  
ID T43646 standard; DNA: 54 BP.  
AC T43646;  
DE 19-AUG-1997 (first entry)  
DT HIV-1 matrix protein p17 gene fragment RYO.  
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
KW detection; maternal transmission; hybridisation assay; immunoassay; ss.  
OS Human immunodeficiency virus type 1.  
PN EP-743364-A2.  
PD 20-NOV-1996.  
PF 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R. Roques P;  
DR WPI: 96-507733/51.

PT Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal  
PT transmission of HIV-1 infection  
PS Claim 3: Page 24; 46pp; French.

CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, GTA, CTA, GTC or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAC, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-fetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 32 A; 5 C; 12 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 ttctaccgtgattgctgcacatc 28  
||||| ||| ||||| |||  
DB 47 TTCTTACTTTTGTGTTGCTCTCTCTCT 21

RESULT 8  
T43647/c  
ID T43647 standard; DNA: 54 BP.  
AC T43647;  
DE 19-AUG-1997 (first entry)  
DT HIV-1 matrix protein p17 gene fragment FLO.  
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
KW detection; maternal transmission; hybridisation assay; immunoassay; ss.  
OS Human immunodeficiency virus type 1.  
PN EP-743364-A2.  
PD 20-NOV-1996.  
PF 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R. Roques P;  
DR WPI: 96-507733/51.

PT Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal  
PT transmission of HIV-1 infection  
PS Claim 3: Page 24; 46pp; French.

CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, GTA, CTA, GTC or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAC, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.

CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-fetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 30 A; 5 C; 14 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 ttctaccgttgattgctgcatcct 28  
||||| ||| ||||| |||  
Db 47 TTCTTACTTTGTGTTGGCTCTCCTCT 21

RESULT 9  
T43648/C  
ID T43648 standard; DNA: 54 BP.  
AC T43648;  
DE 19-AUG-1997 (first entry)  
KW HIV-1 matrix protein p17 gene fragment 4501.  
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
KW detection; maternal transmission; hybridisation assay; immunoassay; ss.  
OS Human immunodeficiency virus type 1.  
FH Key Location/Qualifiers  
FT mat\_peptide 1..54  
FT /tag= a  
FT /product= 4501\_peptide  
PN EP-743364-A2.  
PD 20-NOV-1996.  
PF 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R, Roques P;  
DR WPI: 96-507733/51.  
DR P-PSDB; W06611.  
PT Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal  
PT transmission of HIV-1 infection  
PS Claim 3; Page 24; 46pp; French.  
CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, CTA, GTA, GTG or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-fetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 30 A; 5 C; 14 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 ttctaccgttgattgctgcatcct 28  
||||| ||| ||||| |||  
Db 47 TTCTTACTTTGTGTTGGCTCTCCTCT 21

RESULT 10  
T43649/C  
ID T43649 standard; DNA: 54 BP.  
AC T43649;  
DE 19-AUG-1997 (first entry)  
KW HIV-1 matrix protein p17 gene fragment pAL.  
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;

KW detection; maternal transmission; hybridisation assay; immunoassay; ss.  
OS Human immunodeficiency virus type 1.

PN EP-743364-A2.  
PD 20-NOV-1996.  
PF 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R, Roques P;  
DR WPI: 96-507733/51.

PT Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal  
PT transmission of HIV-1 infection  
PS Claim 3; Page 25; 46pp; French.  
CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, CTA, GTA, GTG or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-fetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 31 A; 4 C; 14 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 ttctaccgttgattgctgcatcct 28  
||||| ||| ||||| |||  
Db 47 TTCTTACTTTGTGTTGGCTCTCCTCT 21

RESULT 11  
T43650/C  
ID T43650 standard; DNA: 54 BP.  
AC T43650;  
DE 19-AUG-1997 (first entry)  
KW HIV-1 matrix protein p17 gene fragment HAR.  
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
KW detection; maternal transmission; hybridisation assay; immunoassay; ss.  
OS Human immunodeficiency virus type 1.  
FH Key Location/Qualifiers  
FT mat\_peptide 1..54  
FT /tag= a  
FT /product= HAR\_peptide  
PN EP-743364-A2.  
PD 20-NOV-1996.  
PF 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R, Roques P;  
DR WPI: 96-507733/51.  
DR P-PSDB; W06612.

PT Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal  
PT transmission of HIV-1 infection  
PS Claim 3; Page 25; 46pp; French.  
CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, CTA, GTA, GTG or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining

CC and assessing the risk of maternal-foetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 29 A; 5 C; 15 G; 4 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ttctacgttgattgctgcattc 28  
Db 47 TTCTACTTTGTTGCTTCCTCCT 21

RESULT 12  
T43652/C  
ID T43652 standard; DNA; 54 BP.  
AC T43652;

DE 19-AUG-1997 (first entry)  
DE HIV-1 matrix protein p17 gene fragment CHET.  
KM Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
KW detection; maternal transmission; hybridisation assay; immunoassay; ss.  
OS Human immunodeficiency virus type 1.  
PN EP-743364-A2.

PD 20-NOV-1996.  
PF 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.

PI Narwa R. Roques P;  
DR WPI: 96-507733/51.

PT Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal  
PT transmission of HIV-1 infection  
PS Claim 3; Page 26; 46pp; French.

CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, GTA, CTA, GAG, GAC, GAT or AGA; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-foetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 30 A; 4 C; 15 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ttctacgttgattgctgcattc 28  
Db 47 TTCTACTTTGTTGCTTCCTCCT 21

RESULT 13  
T43654/C  
ID T43654 standard; DNA; 54 BP.  
AC T43654;

DE 19-AUG-1997 (first entry)  
DE HIV-1 matrix protein p17 gene fragment SIM.  
KM Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
KW detection; maternal transmission; hybridisation assay; immunoassay; ss.  
OS Human immunodeficiency virus type 1.  
PN Location/Qualifiers  
FT Key 1.54  
FT mat\_peptide  
FT /tag= a  
FT /product= SIM\_peptide

PN EP-743364-A2.  
PD 20-NOV-1996.  
PF 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.

PI Narwa R. Roques P;  
DR WPI: 96-507733/51.  
DR P-PSDB; W06614.  
PT Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal  
PT transmission of HIV-1 infection  
PS Claim 3; Page 26; 46pp; French.

CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, GTA, CTA, GAG, GAC, GAT or AGA; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-foetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 30 A; 6 C; 14 G; 4 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ttctacgttgattgctgcattc 28  
Db 47 TTCTACTTTGTTGCTTCCTCCT 21

RESULT 14  
T43655/C  
ID T43655 standard; DNA; 54 BP.  
AC T43655;

DE 19-AUG-1997 (first entry)  
DE HIV-1 matrix protein p17 gene fragment MOE.  
KM Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
KW detection; maternal transmission; hybridisation assay; immunoassay; ss.  
OS Human immunodeficiency virus type 1.  
PN Location/Qualifiers  
FT Key 1.54  
FT mat\_peptide  
FT /tag= a  
FT /product= MOE\_peptide

PD 20-NOV-1996.  
PF 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R. Roques P;  
DR WPI: 96-507733/51.  
DR P-PSDB; W06615.  
PT Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal  
PT transmission of HIV-1 infection  
PS Claim 3; Page 26; 46pp; French.  
CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, GTA, CTA, GAG, GAC, GAT or AGA; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-foetal transmission of HIV-1, using

CC standard hybridisation or immuno assays. The presence of such sequences  
 CC in maternal blood is strongly correlated with transmission of infection.  
 SQ Sequence 54 BP; 30 A; 6 C; 13 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
 Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
 Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 2 ttctacgcgttgattgctgcacatc 28  
 ||||| ||| ||||| |||  
 Db 47 TTCTTACTTTTGTGTGCTCTCTCT 21

## RESULT 15

T43657/c  
 ID T43657 standard; DNA; 54 BP.  
 AC T43657;  
 DT 19-AUG-1997 (first entry)  
 DE HIV-1 matrix protein p17 gene fragment 4541.  
 KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
 KW detection; maternal transmission; hybridisation assay; immunoassay; ss.  
 OS Human immunodeficiency virus type 1.  
 PN EP-743364-A2.  
 PD 20-NOV-1996.  
 PR 17-MAY-1996; 401084.  
 PR 18-MAY-1995; FR-005914.  
 PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
 PI Narwa R, Rognes P;  
 DR WPI; 96-507733/51.  
 PT Human immunodeficiency virus p17 gene fragments, derived proteins  
 PT and antibodies - useful for assessing the risk of maternal  
 PT transmission of HIV-1 infection  
 PS Claim 3; Page 27; 46pp; French.  
 CC This sequence is a specifically claimed example of 21-90 nucleotide  
 CC long nucleic acid fragments, derived from the gene encoding part of the  
 CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
 CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
 CC Y6 = ATA, TTA, CTA, GTA, GTC or ATG; Y7 = GAG or GAA; Y8 = GAA or  
 CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAT or AGA; Y10 = CAA, CAG or  
 CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
 CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
 CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
 CC The new nucleic acid fragments can be used as reagents for determining  
 CC and assessing the risk of maternal-foetal transmission of HIV-1, using  
 CC standard hybridisation or immuno assays. The presence of such sequences  
 CC in maternal blood is strongly correlated with transmission of infection.  
 SQ Sequence 54 BP; 30 A; 4 C; 13 G; 5 T;

Query Match 56.4%; Score 15.8; DB 1; Length 54;  
 Best Local Similarity 74.1%; Pred. No. 2.1e+02;  
 Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 2 ttctacgcgttgattgctgcacatc 28  
 ||||| ||| ||||| |||  
 Db 47 TTCTTACTTTTGTGTGCTCTCTCT 21

Search completed: June 4, 2000, 16:24:08  
 Job time: 28904 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 13:53:12 ; Search time 4521.53 Seconds  
(without alignments)  
25.100 Million cell updates/sec

Title: US-09-164-714-10

Perfect score: 28

Sequence: 1 gtctacgctgattgtctgcatct 28

Scoring table:

IDENTITY\_NDC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 segs, 2026611650 residues

Total number of hits satisfying chosen parameters: 119968

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: gb\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result

Query

SUMMARIES

No.	Score	Match	Length	DB	ID	Description
C 1	15.2	54.3	62	38	AA746404	AA746404 nm62g01.s
C 2	15.2	54.3	63	38	AA746400	AA746400 nm62c07.s
C 3	15.2	54.3	72	30	AA244996	AA244996 nm29d07.r
C 4	15.2	54.3	75	38	AA746399	AA746399 nm62c05.s
C 5	15.2	54.3	75	38	AA746402	AA746402 nm62c11.s
C 6	15.2	54.3	83	37	AA681983	AA681983 vv76c07.r
C 7	15.2	54.3	89	20	D11978	D11978 HUM0512E03
C 8	15.2	54.3	94	20	D12000	D12000 HUM0513A08
C 9	15.2	53.6	84	83	AF088162	AF088162 Homo sapi
C 10	14.8	52.9	81	40	AA930201	AA930201 vs71e03.r
C 11	14.6	52.1	90	47	AI536907	AI536907 col4a09.x
C 12	14.4	51.4	60	29	AA161708	AA161708 M8AFCE9F0
C 13	14.4	51.4	82	64	AA057193	AA057193 ca02a05.y
C 14	14.4	51.4	87	40	AA916294	AA916294 on22b10.s
C 15	14.4	51.4	94	60	AI784864	AI784864 SRAMC30
C 16	14.4	51.4	99	42	AI082999	AI082999 SRAMC318
C 17	14.2	50.7	43	37	AA682184	AA682184 ah51h12.s
C 18	14.2	50.7	85	40	AA910924	AA910924 ok85e02.s
C 19	14.2	50.7	92	24	H99072	H99072 yv90h05.s1
C 20	14	50.0	52	48	AI584790	AI584790 fb83d08.y
C 21	14	50.0	73	33	AA398108	AA398108 zt58d08.s
C 22	14	50.0	82	62	AI906839	AI906839 RC-B1126-
C 23	13.8	49.3	29	20	TI17517	TI17517 gsr m46 The
C 24	13.8	49.3	80	30	AA249355	AA249355 j1284.seq
C 25	13.8	49.3	85	30	AA231275	AA231275 mw32c03.r
C 26	13.8	49.3	100	23	HI9067	HI9067 ym51B08.r1
C 27	13.6	48.6	64	81	AA409557	AA409557 SMOVC3AM
C 28	13.6	48.6	67	44	AI286661	AI286661 ub95902.r
C 29	13.6	48.6	76	35	C58767	C58767 C58767 Yuj1
C 30	13.6	48.6	76	62	AI915855	AI915855 w95a07.x
C 31	13.6	48.6	78	30	AA230217	AA230217 nc13f02.r
C 32	13.6	48.6	88	63	AI960312	AI960312 sc81g04.y
C 33	13.6	48.6	91	79	AA286463	AA286463 lgi 333.B
C 34	13.6	48.6	92	28	C20817	C20817 HDMS000486
C 35	13.6	48.6	94	31	AA291480	AA291480 zt40D04.s
C 36	13.6	48.6	96	46	AI365056	AI365056 qtl2b12.x
C 37	13.6	48.6	97	103	AQ248664	AQ248664 F13111-Sp
C 38	13.6	48.6	99	24	H74881	H74881 226 Deletio
C 39	13.6	48.6	100	34	AA509111	AA509111 M8AFCE9F0
C 40	13.4	47.9	46	94	AA073659	AA073659 EP(2)2525
C 41	13.4	47.9	55	69	AA102242	AA102242 sds5b11.y
C 42	13.4	47.9	58	41	AA008799	AA008799 A0008799
C 43	13.4	47.9	59	63	AI988246	AI988246 sc98a02.y
C 44	13.4	47.9	64	37	AA715443	AA715443 nv53b08.r
C 45	13.4	47.9	67	39	AA905978	AA905978 oj89c03.s

## ALIGNMENTS

RESULT 1  
 LOCUS AA746404 62 bp mRNA EST 27-JAN-1998  
 DEFINITION nm62g01.s1 NCI\_CGAP\_GCB1 Homo sapiens CDNA clone IMAGE:1251169 3',  
 mRNA sequence.  
 ACCESSION AA746404  
 VERSION AA746404.1 GI:2786390  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 62)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,

Ph.D., Gerald Marti, M.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaído, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bdnp/image/image.html](http://www-bio.llnl.gov/bdnp/image/image.html)  
 Insert length: 471 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 57.  
 Location/Qualifiers  
 1. 62  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1251169"  
 /clone\_lib="NCI\_CGAP\_GCB1"  
 /tissue\_type="germinal center B cell"  
 /lab\_host="DH10B"  
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified  
 polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was prepared from human tonsillar cells enriched for  
 germinal center B cells by flow sorting (CD20+, IgD-),  
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
 primed with a Not I - oligo(dT) primer  
 [5'-TGTTACCAATCGAAGGAGGAGGCGCCCTCAATTTTTTTTTTTT-  
 3']. Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pRT3D vector. Library  
 went through one round of normalization, and was  
 constructed by Bento Soares and M. Fatima Bonaído."

BASE COUNT 25 a 11 c 17 g 9 t

Query Match 54.3%; Score 15.2; DB 38; Length 62;  
 Best Local Similarity 71.4%; Pred. No. 7.3e+03;  
 Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Oy 1 gttctaccgttattcttgcattc 28  
 Db 42 gttattgaccttgagttcttcgcatc 15

RESULT 2  
 LOCUS AA746400 63 bp mRNA EST 27-JAN-1998  
 DEFINITION nm62c07.s1 NCI\_CGAP\_GCB1 Homo sapiens CDNA clone IMAGE:1251180 3',  
 mRNA sequence.  
 ACCESSION AA746400  
 VERSION AA746400.1 GI:2786386  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 63)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,  
 Ph.D., Gerald Marti, M.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaído, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be



found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/bbrp/image/image.html

Insert Length: 648 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 63.

FEATURES  
source Location/Qualifiers

1. 63  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1251180"  
/clone\_lib="NCI\_CGAP\_GCB1"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD-),  
provided by Dr. Louis M. Staudt (NCI). Dr. David Allman  
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
primed with a Not I - oligo(dT) primer  
[5'-GGTACCAATCTGAAGGAGCGCGCCATCTTTTCTTTT-  
3']. Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 26 a 10 c 18 g 9 t  
ORIGIN

Query Match 54.3%; Score 15.2; DB 38; Length 63;  
Best Local Similarity 71.4%; Pred. No. 7.3e+03;  
Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 gttctaccgttgattgctgcatct 28  
|| ||| ||||| || |||||  
DB 42 gTATTAGCTTGAGTTCTCTGCATCT 15

RESULT 3 AA244996 72 bp mRNA EST 12-MAR-1997  
LOCUS AA244996/c mw23d07.r1 Guaywoodford Beller mouse kidney day 0 Mus musculus cDNA  
DEFINITION Clone IMAGE:656461 5', mRNA sequence.  
ACCESSION AA244996  
VERSION AA244996.1 GI:1875731  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 72)  
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisler, S., Kucab, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellinger, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HHMI Mouse EST Project  
Unpublished (1996)  
On Dec 30, 1996 this sequence version replaced gi:1530953.  
Contact: Marras M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:402309  
Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 57.

FEATURES  
source Location/Qualifiers

1. 72  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:656461"  
/clone\_lib="Guaywoodford Beller mouse kidney day 0"  
/tissue\_type="kidney"  
/dev\_stage="newborn (day 0)"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: kidney; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR vector;  
-5' adaptor sequence: 5' GAATTCGGACGAG 3' -3' adaptor  
sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3' Library provided  
Lisa Guay-Woodford."

BASE COUNT 33 a 11 c 14 g 13 t 1 others  
ORIGIN

Query Match 54.3%; Score 15.2; DB 30; Length 72;  
Best Local Similarity 71.4%; Pred. No. 7.4e+03;  
Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 gttctaccgttgattgctgcatct 28  
||||||| ||||| |||||  
DB 72 gTCTTACTTTTATTTAGCTTCATCT 45

RESULT 4 AA746399 75 bp mRNA EST 27-JAN-1998  
LOCUS AA746399/c nw62c05.s1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:1251176 3',  
DEFINITION mRNA sequence.  
ACCESSION AA746399  
VERSION AA746399.1 GI:2786385  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 75)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,  
Ph.D., Gerald Marti, M.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/bbrp/image/image.html

Insert Length: 363 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 63.

FEATURES  
source Location/Qualifiers

1. 75  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1251176"  
/clone\_lib="NCI\_CGAP\_GCB1"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for



Matches	20;	Conservative	0;	Mismatches	8;	Indels	0;	Gaps	0;
QY	1	gtcttaccgtgattgctgcgcatct	28						
Db	49	gtcttactcttttaattactcttttcattct	22						
RESULT	7								
LOCUS	D11978	89 bp	mRNA	EST	02-DEC-1992				
DEFINITION	HUM0S12E03 liver HepG2 cell line. Homo sapiens cDNA clone s12e03,								
ACCESSION	D11978								
VERSION	D11978.1	GI:2148809							
KEYWORDS	EST.								
SOURCE	human.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;								
AUTHORS	Eutheria; Primates; Catarrhini; Homiidae; Homo.								
TITLE	1 (bases 1 to 89)								
JOURNAL	Okubo,K., Hori,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y.								
MEDLINE	and Matsubara,K.								
COMMENT	large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression								
FEATURES	Nature Genet. 2, 173-179 (1992)								
source	Contact: Kousaku Okubo, Naohiro Hori, Ryo Matoba, Toshiyuki Niiyama, Atsushi Fukushima, Yoko Kojima & Kenichi Matsubara Institute for Molecular and Cellular Biology Osaka University 1-3 Yamada-Oka,Suita,Osaka 565,Japan.								
BASE COUNT	14 a 21 c 22 g 32 t								
ORIGIN	1. .89								
	/organism="Homo sapiens"								
	/db_xref="GDB:D058180E"								
	/db_xref="taxon:9606"								
	/clone="s12e03"								
	/clone_lib="Liver HepG2 cell line."								
	/lab_host="E.coli"								
	/note="3'-directed regional cDNA library. Cleaved by MBOI and transformed into E.coli."								
Query Match	54.3%; Score 15.2; DB 20; Length 89;								
Best Local Similarity	71.4%; Pred. NO.7.5e+03;								
Matches	20; Conservative	0;	Mismatches	8;	Indels	0;	Gaps	0;	
QY	1	gtcttaccgtgattgctgcgcatct	28						
Db	46	gtcttccctgcttgattgctgcgcatctt	73						
RESULT	8								
LOCUS	D12000	94 bp	mRNA	EST	02-DEC-1992				
DEFINITION	HUM0S13A08 liver HepG2 cell line. Homo sapiens cDNA clone s13a08,								
ACCESSION	D12000								
VERSION	D12000.1	GI:2148823							
KEYWORDS	EST.								
SOURCE	human.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;								
AUTHORS	Eutheria; Primates; Catarrhini; Homiidae; Homo.								
TITLE	1 (bases 1 to 94)								
JOURNAL	Okubo,K., Hori,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y.								
MEDLINE	large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression								
	Nature Genet. 2, 173-179 (1992)								

```

COMMENT      Contact: Kousaku Okubo, Naohito Hori, Ryo Matoba, Toshiyuki Niiyama, Atsushi Fukushima, Yoko Kojima & Kenichi Matsubara Institute for Molecular and Cellular Biology Osaka University  
              1-3 Yamada-oka, Suita, Osaka 565, Japan.  
  
FEATURES     Location/Qualifiers  
             source          1..94  
               /organism="Homo sapiens"  
               /db_xref="GDB:D08194E"  
               /db_xref="taxon:9606"  
               /clone="g13ag08"  
               /clone_lib="Liver HepG2 cell line."  
               /lab_host="E.coli"  
               /note="3"-directed regional cDNA library. Cleaved by MboI and transformed into E.coli."
BASE COUNT   18 a         21 c        21 g       33 t           1 others
ORIGIN
Query Match    54.3%; Score 15.2; DB 20; Length 94;  
Best Local Similarity 71.4%; Pred. No. 7.5e+03;  
Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY            1 gtcttaccgttgattgcctgatc 28  
Db            46 GCCTCCCGTGTGAATGGTCGTGCATT 73  
| | | | |  
RESULT 9  
AF088162      84 bp DNA GSS 10-DEC-1998  
LOCUS Homo sapiens chromosome 11 clone 49D4Ec3 map 11q13, genomic survey sequence.  
ACCESSION AF088162  
VERSION APF088162.1 GI:3982872  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Human  
            Homo sapiens  
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Euarchonta; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 84)  
AUTHORS Gerhart,D.S. and Zhu,S.  
TITLE A transcript map of an 800 kb region on human chromosome 11q13,  
PART OF part of the candidate region for SCA5 and BBS1  
JOURNAL Hum. Genet. (1999) In press  
REFERENCE 2 (bases 1 to 84)  
AUTHORS Gerhard,D.S. and Zhu,S.  
TITLE Direct Submission  
JOURNAL Submitted (31-AUG-1998) Department of Genetics, Washington University School of Medicine, 4566 Scott Avenue, Box 8232, St. Louis, MO 63110, USA  
  
FEATURES     Location/Qualifiers  
             source          1..84  
               /organism="Homo sapiens"  
               /db_xref="taxon:9606"  
               /map="11q13"  
               /clone="49D4Ec3"  
               /chromosome="11"  
BASE COUNT   24 a         17 c        10 g       33 t  
ORIGIN
```

LOCUS AA930201 81 bp mRNA EST 23-APR-1998  
DEFINITION vs1le03.r1 Strataegene mouse skin (#937313) Mus musculus cDNA clone  
IMAGE:1151740 5' similar to gb:U15647\_cds1 Mus musculus (MOUSE),,  
MRNA sequence.  
ACCESSION AA930201  
VERSION AA930201.1 GI:3079794  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS 1 (bases 1 to 81)  
Marras, M., Hille, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Jan 19, 1998 this sequence version replaced gi:2286367.  
Contact: Maria M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:624948  
Seq primer: -28m13 rev1 ET from Amersham.  
FEATURES  
source Location/Qualifiers  
1..81  
/organism="Mus musculus"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1151740"  
/clone\_1lb="Strataegene mouse skin (#937313)"  
/sex="females"  
/tissue\_type="whole skin"  
/dev\_stage="11 weeks old"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: skin; Vector: pBluescript SK-; Site\_1:  
ECORI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. Whole skin from 11 week old C57BL/6 female mice.  
Average insert size: 1.0 kb; Uni-ZAP XR Vector: ~5'  
adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor  
sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'."  
BASE COUNT 35 a 21 c 9 g 16 t  
ORIGIN  
Query Match 52.9%; Score 14.8; DB 40; Length 81;  
Best Local Similarity 73.1%; Pred. No. 1.1e+04;  
Matches 19; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
Qy 3 tcttaccgttgattgctgcacatc 28  
||||| ||||| ||||| ||||| |||||  
Db 66 tcttcagcggttgcttgatgcatctt 41  
RESULT 11  
LOCUS A1536907 90 bp mRNA EST 12-MAY-1999  
DEFINITION t014a09.x1 NCI\_CGAP\_Ut2 Homo sapiens cDNA clone IMAGE:2179000 3',  
MRNA sequence.  
ACCESSION A1536907  
VERSION A1536907.1 GI:4451042  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS 1 (bases 1 to 90)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On May 18, 1998 this sequence version replaced gi:3138691.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www.bio.lnl.gov/bdrp/image/image.html  
Insert Length: 1143 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 85  
POLYA-No.  
FEATURES  
source Location/Qualifiers  
1..90  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2179000"  
/clone\_1lb="NCI\_CGAP Ut2"  
/tissue\_type="moderately-differentiated endometrial  
adenocarcinoma, 3 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: PCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.85 kb. Life Technologies catalog #: 11539-012"  
BASE COUNT 34 a 16 c 13 g 27 t  
ORIGIN  
Query Match 52.1%; Score 14.6; DB 47; Length 90;  
Best Local Similarity 81.0%; Pred. No. 1.4e+04;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 2 tcttaccgttgattgctg 22  
||||| ||||| ||||| ||||| |||||  
Db 37 tttcttaccgttgctgcttg 17  
RESULT 12  
LOCUS A161708 60 bp mRNA EST 17-DEC-1996  
DEFINITION MBARCE9P05r3 Brugia malayi adult female cDNA (SAM96MLM-BNAF) Brugia  
malayi cDNA clone AFCE9F05 5', mRNA sequence.  
ACCESSION A161708  
VERSION A161708.1 GI:1737540  
KEYWORDS EST.  
SOURCE Brugia malayi.  
ORGANISM Brugia malayi  
Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida;  
Filarioidea; Onchocercidae; Brugia.  
REFERENCE  
AUTHORS 1 (bases 1 to 60)  
Blaxter, M.L., Waterfall, M., Daub, J., Lizotte, M., Baron, L. and  
Jones, S.J.  
Genes expressed in adult female Brugia malayi  
Contact: Blaxter ML  
On May 18, 1995 this sequence version replaced gi:811474.  
JOURNAL Unpublished (1996)  
COMMENT Institute of Cell, Animal and Population Biology  
University of Edinburgh  
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9  
3JT, UK.  
Tel: +44 131 650 6760  
Fax: +44 131 670 5450

Email: mark.blaxter@ed.ac.uk  
The ABI trace of this sequence can be viewed at  
http://www.sanger.ac.uk/Brugia/ABC/MBAFCE9F05T3.html  
Seq primer: T3.

## FEATURES

source

Location/Qualifiers

1..60

/organism="Brugia malayi"

/db\_xref="taxon:6279"

/clone="AFCE9F05"

/clone\_1lb="Brugia malayi adult female cDNA

(SAM96MLM-BMAF)"

/sex="female"

/dev\_stage="adult"

/lab\_host="XLI-Blue MRF"

/note="Vector: Lambda Uni-ZAP XR; Site\_1: Eco RI; Site\_2:

Xho I; Lymphatic filarial nematode parasite of humans.

mRNA was prepared from approximately 50 adult females

isolated from the peritoneal cavity of jirds and

converted to double-stranded cDNA using reverse

transcriptase and oligo(dT) followed by RNase H and DNA

pol I. The library has 5 x 10E6 independent recombinants

and the average insert size is ~900bp. The library was

constructed by Michelle Lizotte-Waniewski. The

library is available from Dr.S.A.Williams, email:

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

genome@smith.edu."

Query Match 51.4%; Score 14.4; DB 29; Length 60;  
Best Local Similarity 75.0%; Pred. No. 1.6e+04;  
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 tctaccgtgattgctgcat 26  
||| | ||| ||||| |||  
Db 2 TCTAATTTTGTTCCTGCTGAT 25

RESULT 13  
AM057193 82 bp mRNA EST 29-SEP-1999  
LOCUS ca02a05.y1 C elegans fem3 Q23 S1 Caenorhabditis elegans cDNA 5'  
DEFINITION similar to gb:R13H9.2 (ELEGENS); WP:F09C12.7 CE02603 ;, mRNA  
sequence.  
ACCESSION AM057193  
VERSION AM057193.1 GI:5932832  
KEYWORDS EST.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;  
Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
1 (bases 1 to 82)  
Ward,S., Smith,H., Clifton,S., Marra,M., Hillier,L., Kucaba,T.,  
Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Allen,M.,  
Bowers,Y., Person,B., Swaller,T., Steptoe,M., Gibbons,M.,  
Harvey,N., Riller,E., Jackson,Y., McCann,R., Waterston,R. and  
Wilson,R.

TITLE UofArizona-Washu C. elegans EST project  
JOURNAL Unpublished (1999)  
COMMENT On Dec 20, 1995 this sequence version replaced gi:1135389.  
Contact: Samuel Ward, Ph.D.  
UofArizona-Washu C. elegans EST project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
Contact Harold Smith (hes@u.arizona.edu) for further information  
relating to organism, libraries, or clone availability.  
Seq primer: -40UP from Gibco.

FEATURES  
source 1..82  
/organism="Caenorhabditis elegans"

## BASE COUNT

20 a 22 c 22 g 18 t

## ORIGIN

Query Match 51.4%; Score 14.4; DB 64; Length 82;  
Best Local Similarity 75.0%; Pred. No. 1.6e+04;  
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 5 ttaccgtgattgctgcatc 28  
||| | ||| ||||| |||  
Db 32 TTAACGGCGAATTCCTGCGAGCT 9

RESULT 14  
AA916294 87 bp mRNA EST 17-JUN-1998  
LOCUS on22b10.s1 NCI-CGAP\_Lu5 Homo sapiens cDNA clone IMAGE:1557403 3'  
DEFINITION similar to gb:X61123\_rnal BRG1 PROTEIN (HUMAN);, mRNA sequence.  
ACCESSION AA916294  
VERSION AA916294.1 GI:3055686  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 87)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On May 5, 1995 this sequence version replaced gi:797728.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1350  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA library preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILN at:  
www.bio.llnl.gov/dbp/image/image.html

Trace considered overall poor quality  
Insert Length: 990 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers  
1..87  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1557403"  
/clone\_1lb="NCI-CGAP\_Lu5"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/note="Organ: Lung; Vector: pTZ19-Pac (Pharmacia) with a  
modified polylinker; 1st strand cDNA was prepared from

neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia) and digested with Not I and cloned into the Not I and Eco RI sites of the modified pT773 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "

```

Db      21 TCATATTTTGTTCCTGGTAT 44

Search completed: June  4, 2000, 13:53:12
Job time: 20936 sec

```

Query Match	51.4%;	Score 14.4;	DB 40;	Length 87;
Best Local Similarity	75.0%;	Pred. No. 1.6e+04;		
Matches 18; Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0;

```
QY      5 ttaccgttgattgctggcatct 28
          | | ||| ||| ||||| |
Db     26 TAATCGTTCCTTCTTGGCATT 3
```

RESULT	15
LOCUS	A1784864
DEFINITION	A1784864 94 bp mRNA EST 02-JUL-1999
ACCESSION	SAMMCC30B03K Brugia malayi adult male cDNA (SAM94NL-Bmam) Brugia malayi cDNA clone SAMMCC30B03 5', mRNA sequence.
VERSION	A1784864
KEYWORDS	A1784864.1 GI:5332489
SOURCE	EST
ORGANISM	Brugia malayi. Brugia malayi

REFERENCE	1 (bases 1 to 94)
AUTHORS	Williams, S. A.
TITLE	Genes expressed in adult males of <i>Brugia malayi</i>
JOURNAL	Unpublished (1993)
COMMENT	On Jun 22, 1998 this sequence version replaced gi:3246848.

Email: genome@smith.edu	
Seq primer: pbluescript SK.	
location/Qualifiers	
1. .94	

source

1. 94

/organism="Brugia\_malay1"

/strain="TR3 Labs"

/db\_xref="taxon:6279"

/clone="SWAMCACC30B03"

/lab\_host="Brugia\_malay1 adult male CDNA (SAM94NL-BimM)"

/lab\_host="XLI-Blue MRF"

/note="Vector: lambda Unizap XR; Site\_1: EcoR I; Site\_2: Xho I; lymphatic filarial nematode parasite of humans. mRNA was prepared from adult males of Brugia malayi isolated from jirds and converted to double stranded cDNA using reverse transcriptase and oligo(dT) followed by Rase H and Dnasep I. The library had 4.6 x 10<sup>6</sup> independent recombinants and average insert size was 800 base pairs. The library was constructed by Noelle Ling. The library is available from Dr. S.A. Williams, email: genome@smith.edu."

19 a 4 c 9 g 62 t

BASE COUNT

ORIGIN

Query Match	51.48;	Score 14.4;	DB 60;	Length 94;
Best Local Similarity	75.08;	Pred. No. 1.7e+04;		
Matches 18; Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0;

Qy 3 tctaccgtgattgcttgcac 26  
||| | ||| ||||| ||| ||

```

Db      21 TCATATTTTGTTCCTGGTAT 44

Search completed: June  4, 2000, 13:53:12
Job time: 20936 sec

```

---

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:09:42 ; Search time 244.64 Seconds  
(without alignments)  
14.877 Million cell updates/sec

Title: US-09-164-714-10

Perfect score: 28  
Sequence: 1 gtcttaccgttgattgttgatcatc 28

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 375880

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/lna/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/lna/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/lna/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/lna/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/lna/6\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/lna/PCrUS\_COMB.seq:\*  
7: /cgn2\_6/ptodata/2/lna/Backfillseq1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	57.1	79	2	US-08-446-908-8
2	16	57.1	79	2	US-08-231-205A-8
3	16	57.1	79	4	US-08-871-161-8
4	15.8	56.4	54	4	US-08-649-991-10
5	15.8	56.4	54	4	US-08-649-991-11
6	15.8	56.4	54	4	US-08-649-991-12
7	15.8	56.4	54	4	US-08-649-991-13
8	15.8	56.4	54	4	US-08-649-991-14
9	15.8	56.4	54	4	US-08-649-991-15
10	15.8	56.4	54	4	US-08-649-991-16
11	15.8	56.4	54	4	US-08-649-991-17
12	15.8	56.4	54	4	US-08-649-991-18
13	15.8	56.4	54	4	US-08-649-991-20
14	15.8	56.4	54	4	US-08-649-991-22
15	15.8	56.4	54	4	US-08-649-991-23
16	15.8	56.4	54	4	US-08-649-991-25
17	15.8	56.4	54	4	US-08-649-991-26
18	15.8	56.4	54	4	US-08-649-991-27
19	15.8	56.4	54	4	US-08-649-991-119
20	15.8	56.4	54	4	US-08-649-991-120
21	15.8	56.4	54	4	US-08-649-991-121
22	15.8	56.4	54	4	US-08-649-991-122
23	15.8	56.4	54	4	US-08-649-991-123
24	15.8	56.4	54	4	US-08-649-991-124
25	15.8	56.4	54	4	US-08-649-991-125
26	15.8	56.4	54	4	US-08-649-991-126
27	15.8	56.4	54	4	US-08-649-991-128

28	15.8	56.4	54	4	US-08-649-991-129	Sequence 129, App
29	15	53.6	31	1	US-08-390-850-43	Sequence 43, App
30	15	53.6	31	2	US-08-435-634-43	Sequence 43, App
31	14.8	52.9	53	3	US-08-811-492-120	Sequence 120, App
32	14.8	52.9	54	4	US-08-649-991-21	Sequence 21, App
33	14.8	52.9	54	4	US-08-649-991-24	Sequence 24, App
34	14.8	52.9	54	4	US-08-649-991-31	Sequence 31, App
35	14.8	52.9	54	6	PCr-US94-09700-35	Sequence 35, App
36	14.8	52.9	65	2	US-08-273-594-29	Sequence 29, App
37	14.6	52.1	69	3	US-07-916-098A-26	Sequence 26, App
38	14.6	52.1	69	3	US-07-916-098A-31	Sequence 31, App
39	14.2	50.7	50	4	US-08-850-049-5	Sequence 5, App
40	14.2	50.7	50	4	US-08-050-478-5	Sequence 5, App
41	14.2	50.7	54	4	US-08-649-991-9	Sequence 9, App
42	14.2	50.7	54	4	US-08-649-991-19	Sequence 19, App
43	14.2	50.7	57	7	5514566-17	Patent No. 5514566
44	14.2	50.7	99	2	US-08-463-115-69	Sequence 69, App
45	14.2	50.7	99	2	US-08-465-388-69	Sequence 69, App

## ALIGNMENTS

RESULT 1  
US-08-446-908-8/c  
Sequence 8, Application US/08446908  
Patent No. 5705149  
GENERAL INFORMATION:  
APPLICANT: Namen, Anthony E.  
APPLICANT: Goodwin, Raymond G.  
APPLICANT: Lupton, Stephen D.  
APPLICANT: Mochizuki, Diane Y.  
TITLE OF INVENTION: Interleukin-7 and Antibodies Reactive  
TITLE OF INVENTION: Therewith  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: WA  
COUNTRY: US  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple 7.1  
SOFTWARE: Microsoft Word, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446,908  
FILING DATE: 22-May-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/231,205  
FILING DATE: 21-Apr-1994  
APPLICATION NUMBER: US 07/957,649  
FILING DATE: 06-Oct-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/511,438  
FILING DATE: 13-Apr-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/255,209  
FILING DATE: 07-Oct-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/113,566  
FILING DATE: 26-Oct-1987  
ATTORNEY/AGENT INFORMATION:  
NAME: Seese, Kathryn A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2104-D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
INFORMATION FOR SEQ ID NO: 8:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 79 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 19..60
; US-08-446-908-8

Query Match          57.1%; Score 16; DB 2; Length 79;
Best Local Similarity 79.2%; Pred. No. 50;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 ctaccgtgattctgttcacgc 27
    | |||| | |||| | |||| | ||||
Db 57 CCTACCTGCGATTCTGTCTCATGC 34

RESULT 2
US-08-231-205A-8/c
; Sequence 8, Application US/08231205A
; Patent No. 5714585
; GENERAL INFORMATION:
; APPLICANT: Namen, Anthony E.
; APPLICANT: Goodwin, Raymond G.
; APPLICANT: Lupton, Stephen D.
; APPLICANT: Mochizuki, Diane Y.
; TITLE OF INVENTION: Interleukin-7 and Antibodies Reactive
; TITLE OF INVENTION: Therewith
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: US
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple 7.1
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,205A
; FILING DATE: 21-APR-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/957,649
; FILING DATE: 06-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,438
; FILING DATE: 13-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/255,209
; FILING DATE: 07-OCT-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/113,566
; FILING DATE: 26-OCT-1987
; ATTORNEY/AGENT INFORMATION:
; NAME: Seese, Kathryn A.
; REGISTRATION NUMBER: 32,172
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 79 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

```

; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 19..60
; US-08-231-205A-8

Query Match          57.1%; Score 16; DB 2; Length 79;
Best Local Similarity 79.2%; Pred. No. 50;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 ctaccgtgattctgttcacgc 27
    | |||| | |||| | |||| | ||||
Db 57 CCTACCTGCGATTCTGTCTCATGC 34

RESULT 3
US-08-871-161-8/c
; Sequence 8, Application US/08871161
; Patent No. 5965122
; GENERAL INFORMATION:
; APPLICANT: Namen, Anthony E.
; APPLICANT: Goodwin, Raymond G.
; APPLICANT: Lupton, Stephen D.
; APPLICANT: Mochizuki, Diane Y.
; TITLE OF INVENTION: Interleukin-7 and Antibodies Reactive
; TITLE OF INVENTION: Therewith
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: US
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple 7.1
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871,161
; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/446,908
; FILING DATE: 22-MAY-1995
; APPLICATION NUMBER: US 08/231,205
; FILING DATE: 21-APR-1994
; APPLICATION NUMBER: US 07/957,649
; FILING DATE: 06-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,438
; FILING DATE: 13-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/255,209
; FILING DATE: 07-OCT-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/113,566
; FILING DATE: 26-OCT-1987
; ATTORNEY/AGENT INFORMATION:
; NAME: Seese, Kathryn A.
; REGISTRATION NUMBER: 32,172
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 79 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```



CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036-5869  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/649,991  
FILING DATE: 17-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9505914  
FILING DATE: 18-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Adler, Reid G.  
REGISTRATION NUMBER: 30,988  
REFERENCE/DOCKET NUMBER: ORES-5003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-467-7000  
TELEFAX: 202-467-7176  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-649-991-12

Query Match 56.4%; Score 15.8; DB 4; Length 54;  
Best Local Similarity 74.1%; Pred. No. 57;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 ttctaccgttgattgctggcatct 28  
||||| ||| ||||| | |||  
DB 47 TTCCTACTTTGTTTGCTCTCTCT 21

RESULT 7  
US-08-649-991-13/c  
Sequence 13, Application US/08649991  
Patent No. 5919462  
GENERAL INFORMATION:  
APPLICANT: Narva, Remy  
TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE  
TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR  
TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF  
TITLE OF INVENTION: MATERNOPFETAL TRANSMISSION OF HIV-1  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP  
STREET: 1800 M Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036-5869  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/649,991  
FILING DATE: 17-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9505914  
FILING DATE: 18-MAY-1995

ATTORNEY/AGENT INFORMATION:  
NAME: Adler, Reid G.  
REGISTRATION NUMBER: 30,988  
REFERENCE/DOCKET NUMBER: ORES-5003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-467-7000  
TELEFAX: 202-467-7176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-649-991-13

Query Match 56.4%; Score 15.8; DB 4; Length 54;  
Best Local Similarity 74.1%; Pred. No. 57;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 ttctaccgttgattgctggcatct 28  
||||| ||| ||||| | |||  
DB 47 TTCCTACTTTGTTTGCTCTCTCT 21

RESULT 8  
US-08-649-991-14/c  
Sequence 14, Application US/08649991  
Patent No. 5919462  
GENERAL INFORMATION:  
APPLICANT: Narva, Remy  
TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE  
TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR  
TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF  
TITLE OF INVENTION: MATERNOPFETAL TRANSMISSION OF HIV-1  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP  
STREET: 1800 M Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036-5869  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/649,991  
FILING DATE: 17-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9505914  
FILING DATE: 18-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Adler, Reid G.  
REGISTRATION NUMBER: 30,988  
REFERENCE/DOCKET NUMBER: ORES-5003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-467-7000  
TELEFAX: 202-467-7176  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-649-991-14

Query Match 56.4% Score 15.8; DB 4; Length 54;  
Best Local Similarity 74.1%; Pred. No. 57;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 ttctaccgttgattgctgcatc 28  
||||| ||| ||||| |||  
DB 47 TTCTTACTTTTGTTCCTCTCTCT 21

RESULT 9  
US-08-649-991-15/c  
; Sequence 15, Application US/08649991  
; Patent No. 5919462

GENERAL INFORMATION:  
APPLICANT: Narwa, Remy  
TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE  
TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR  
TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF  
TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP  
STREET: 1800 M Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036-5869  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/649,991  
FILING DATE: 17-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9505914  
FILING DATE: 18-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Adler, Reid G.  
REGISTRATION NUMBER: 30,988  
REFERENCE/DOCKET NUMBER: ORES-5003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-467-7000  
TELEFAX: 202-467-7176  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-649-991-15

Query Match 56.4% Score 15.8; DB 4; Length 54;  
Best Local Similarity 74.1%; Pred. No. 57;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 ttctaccgttgattgctgcatc 28  
||||| ||| ||||| |||  
DB 47 TTCTTACTTTTGTTCCTCTCTCT 21

RESULT 10  
US-08-649-991-16/c  
; Sequence 16, Application US/08649991  
; Patent No. 5919462  
; GENERAL INFORMATION:  
APPLICANT: Narwa, Remy

APPLICANT: Roques, Pierre  
TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE  
TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR  
TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF  
TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP  
STREET: 1800 M Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036-5869

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/649,991  
FILING DATE: 17-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9505914  
FILING DATE: 18-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Adler, Reid G.  
REGISTRATION NUMBER: 30,988  
REFERENCE/DOCKET NUMBER: ORES-5003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-467-7000  
TELEFAX: 202-467-7176  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-649-991-16

Query Match 56.4% Score 15.8; DB 4; Length 54;  
Best Local Similarity 74.1%; Pred. No. 57;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 ttctaccgttgattgctgcatc 28  
||||| ||| ||||| |||  
DB 47 TTCTTACTTTTGTTCCTCTCTCT 21

RESULT 11  
US-08-649-991-17/c  
; Sequence 17, Application US/08649991  
; Patent No. 5919462  
; GENERAL INFORMATION:  
APPLICANT: Narwa, Remy  
TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE  
TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR  
TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF  
TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP  
STREET: 1800 M Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036-5869  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/649,991  
FILING DATE: 17-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9505914  
FILING DATE: 18-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Adler, Reid G.  
REGISTRATION NUMBER: 30,988  
REFERENCE/DOCKET NUMBER: ORES-5003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-467-7176  
TELEFAX: 202-467-7000  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-649-991-17

Query Match 56.4%; Score 15.8; DB 4; Length 54;  
Best Local Similarity 74.1%; Pred. No. 57;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ttctacgtgattgctgcatcct 28  
||||| ||| ||||| | |||  
Db 47 ttctacttttcttgcctcctcct 21

RESULT 12  
US-08-649-991-18/c  
Sequence 18, Application US/08649991  
Patent No. 5919462  
GENERAL INFORMATION:  
APPLICANT: Naiva, Remy  
TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE  
TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR  
TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP  
STREET: 1800 M Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036-5869  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/649,991  
FILING DATE: 17-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9505914  
FILING DATE: 18-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Adler, Reid G.  
REGISTRATION NUMBER: 30,988  
REFERENCE/DOCKET NUMBER: ORES-5003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-467-7000  
TELEFAX: 202-467-7176

INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-649-991-18

Query Match 56.4%; Score 15.8; DB 4; Length 54;  
Best Local Similarity 74.1%; Pred. No. 57;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ttctacgtgattgctgcatcct 28  
||||| ||| ||||| | |||  
Db 47 ttctacttttcttgcctcctcct 21

RESULT 13  
US-08-649-991-20/c  
Sequence 20, Application US/08649991  
Patent No. 5919462  
GENERAL INFORMATION:  
APPLICANT: Naiva, Remy  
TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE  
TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR  
TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP  
STREET: 1800 M Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036-5869  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/649,991  
FILING DATE: 17-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9505914  
FILING DATE: 18-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Adler, Reid G.  
REGISTRATION NUMBER: 30,988  
REFERENCE/DOCKET NUMBER: ORES-5003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-467-7176  
TELEFAX: 202-467-7000  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-649-991-20

Query Match 56.4%; Score 15.8; DB 4; Length 54;  
Best Local Similarity 74.1%; Pred. No. 57;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 ttctacgtgattgctgcatcct 28  
||||| ||| ||||| | |||

Db 47 TTTCTACTTTTGTGCTCTCTCTCT 21

RESULT 14  
US-08-649-991-22/c

; Sequence 22, Application US/08649991

; Patent No. 5919462

; GENERAL INFORMATION:

; APPLICANT: Naiva, Remy

; APPLICANT: Rogues, Pierre

; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE

; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR

; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF

; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1

; NUMBER OF SEQUENCES: 130

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP

; STREET: 1800 M Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20036-5869

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/649,991

; FILING DATE: 17-MAY-1996

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: FR 9505914

; FILING DATE: 18-MAY-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Adler, Reid G.

; REGISTRATION NUMBER: 30,988

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-467-7000

; TELEFAX: 202-467-7176

; INFORMATION FOR SEQ ID NO: 22:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 54 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; US-08-649-991-22

Query Match

Best Local Similarity 56.4%; Score 15.8; DB 4; Length 54;

Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 ttctacgctgattgctgcctc 28

Db 47 TTTCTACTTTTGTGCTCTCTCTCT 21

RESULT 15

US-08-649-991-23/c

; Sequence 23, Application US/08649991

; Patent No. 5919462

; GENERAL INFORMATION:

; APPLICANT: Naiva, Remy

; APPLICANT: Rogues, Pierre

; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE

; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR

; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF

; NUMBER OF SEQUENCES: 130

; CORRESPONDENCE ADDRESSES:

ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP  
STREET: 1800 M Street, N.W.  
CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20036-5869

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/649,991

FILING DATE: 17-MAY-1996

CLASSIFICATION: 435

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: FR 9505914

FILING DATE: 18-MAY-1995

ATTORNEY/AGENT INFORMATION:

NAME: Adler, Reid G.

REGISTRATION NUMBER: 30,988

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-467-7000

TELEFAX: 202-467-7176

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 54 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

; US-08-649-991-23

Query Match

Best Local Similarity 56.4%; Score 15.8; DB 4; Length 54;

Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 ttctacgctgattgctgcctc 28

Db 47 TTTCTACTTTTGTGCTCTCTCTCT 21

Search completed: June 4, 2000, 16:09:42

Job time: 28061 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:05:33 ; Search time 1236.38 Seconds

(without alignments)  
-22.031 Million cell updates/sec

Title: US-09-164-714-11

Perfect score: 28  
Sequence: 1 ctgaccaatttggcaccacaataagag 28

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 356616

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pr1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vl:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_or:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vl:\*  
32: gb\_htg1:\*  
33: gb\_htg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba2:\*  
37: em\_ba1:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pr4:\*  
41: gb\_htg3:\*  
42: gb\_htg4:\*  
43: gb\_htg5:\*  
44: gb\_htg6:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	57.1	39	5	AR012248
2	16	57.1	39	5	I14946
3	16	57.1	39	5	I73666
4	15	53.6	34	5	A39689
5	15	53.6	34	5	A48961
6	14.8	52.9	75	40	AF064903
7	14.8	52.9	92	10	HUMPHOS15
8	14.2	50.7	53	4	FR08001
9	14.2	50.7	60	8	AF101150
10	14.2	50.7	72	5	AR040932
11	14.2	50.7	72	5	AR051839
12	14.2	50.7	72	5	I36229
13	14.2	50.7	90	9	GORINSORFA
14	14.2	50.7	99	34	DD1DKE
15	14	50.0	27	5	E14994
16	14	50.0	36	5	A78745
17	14	50.0	36	5	AR014718
18	13.8	49.3	35	5	AR055776
19	13.8	49.3	35	5	AR055777
20	13.8	49.3	51	12	U92178
21	13.6	48.6	26	5	AR014478
22	13.6	48.6	59	10	S59972
23	13.6	48.6	60	10	S72312
24	13.6	48.6	81	3	PIGBROG
25	13.6	48.6	99	12	MUSLRP03
26	13.4	47.9	44	5	AR048165
27	13.4	47.9	77	13	R1C1254A
28	13.4	47.9	87	10	S71607
29	13.2	47.1	23	5	AR063157
30	13.2	47.1	29	5	AR003150
31	13.2	47.1	29	5	AR003243
32	13.2	47.1	29	5	AR009108
33	13.2	47.1	29	5	AR011395
34	13.2	47.1	29	5	AR028081
35	13.2	47.1	29	5	AR052797
36	13.2	47.1	29	5	I18033
37	13.2	47.1	43	5	A05133
38	13.2	47.1	50	9	HPBHUMIN2
39	13.2	47.1	72	5	I00803
40	13.2	47.1	77	5	A00281
41	13.2	47.1	77	5	A16012
42	13.2	47.1	93	12	MMG86712
43	13	46.4	21	5	AR024394
44	13	46.4	21	5	I71186
45	13	46.4	21	5	I83745

## ALIGNMENTS

RESULT 1  
AR012248 AR012248 39 bp DNA PAT 04-DEC-1998  
LOCUS Sequence 32 from patent US 5763254.  
DEFINITION AR012248  
ACCESSION AR012248  
VERSION AR012248.1 GI:3970238  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE Unclassified.  
AUTHORS 1 (bases 1 to 39)  
TITLE Woldike,H.,Fabricius, Hagen,F., Hjort,C.,Malland and Hastrup,S.  
JOURNAL Enzyme capable of degrading cellulose or hemicellulose  
FEATURES Patent: US 5763254-A 32 09-JUN-1998;  
source Location/Qualifiers  
1. .39  
/organism="unknown"  
BASE COUNT 9 a 16 c 8 g 6 t  
ORIGIN

Query Match 57.1%; Score 16; DB 5; Length 39;  
Best Local Similarity 79.2%; Pred. NO. 2.5e+03;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 gaccatttggcaccacatagg 27  
||||| ||| | ||| |  
Db 11 GACCAATTCTGCGCCACACATTGG 34

RESULT 2  
I14946 I14946 39 bp DNA PAT 02-APR-1996  
LOCUS Sequence 32 from patent US 5457046.  
DEFINITION I14946  
ACCESSION I14946  
VERSION I14946.1 GI:1249854  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
AUTHORS 1 (bases 1 to 39)  
TITLE Woldike,H.,F., Hagen,F., Hjort,C. and Hastrup,S.  
JOURNAL Enzyme capable of degrading cellulose or hemicellulose  
FEATURES Patent: US 5457046-A 32 10-OCT-1995;  
source Location/Qualifiers  
1. .39  
/organism="unknown"  
BASE COUNT 9 a 16 c 8 g 6 t  
ORIGIN

Query Match 57.1%; Score 16; DB 5; Length 39;  
Best Local Similarity 79.2%; Pred. NO. 2.5e+03;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 gaccatttggcaccacatagg 27  
||||| ||| | ||| |  
Db 11 GACCAATTCTGCGCCACACATTGG 34

RESULT 3  
I73666 I73666 39 bp DNA PAT 03-APR-1998  
LOCUS Sequence 32 from patent US 5686593.  
DEFINITION I73666  
ACCESSION I73666  
VERSION I73666.1 GI:3009807  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
AUTHORS 1 (bases 1 to 39)  
TITLE Woldike,H.,Fabricius, Hagen,F., Hjort,C.,Malland and Hastrup,S.

TITLE Enzyme capable of degrading cellulose or hemicellulose  
JOURNAL Patent: US 5686593-A 32 11-NOV-1997;  
FEATURES Location/Qualifiers  
source 1. .39  
/organism="unknown"  
BASE COUNT 9 a 16 c 8 g 6 t  
ORIGIN

Query Match 57.1%; Score 16; DB 5; Length 39;  
Best Local Similarity 79.2%; Pred. NO. 2.5e+03;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 gaccatttggcaccacatagg 27  
||||| ||| | ||| |  
Db 11 GACCAATTCTGCGCCACACATTGG 34

RESULT 4  
A39689/c A39689 34 bp DNA PAT 05-MAR-1997  
LOCUS Sequence 2 from Patent W09418327.  
DEFINITION A39689  
ACCESSION A39689  
VERSION A39689.1 GI:2295950  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 34)  
AUTHORS Burnham,M.K., Chopra,I., Critchley,I.A. and Knowles,D.J.  
TITLE FIBRONECTIN BINDING PROTEIN; MONOCLONAL ANTIBODY AND THEIR USE IN PREVENTING BACTERIAL ADHESION  
JOURNAL Patent: WO 9418327-A 2 18-AUG-1994;  
COMMENT SMITHKLINE BEECHAM PLC (GB)  
Other publication AU 5975994 940829  
Other publication CN 1119026 960320  
Other publication CA 2155413 940818  
Other publication ZA 9400728 941107  
Other publication JP 85068107 960723.  
source Location/Qualifiers  
1. .34  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 5 a 6 c 14 g 9 t  
ORIGIN

Query Match 53.6%; Score 15; DB 5; Length 34;  
Best Local Similarity 78.3%; Pred. NO. 7.2e+03;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 6 ccaatttggcaccacatagg 28  
||||| ||| ||| ||| ||| |||  
Db 29 CCAATCTGCGCCACACGTAAGG 7

RESULT 5  
A48961/c A48961 34 bp DNA PAT 07-MAR-1997  
LOCUS Sequence 8 from Patent W09604380.  
DEFINITION A48961  
ACCESSION A48961  
VERSION A48961.1 GI:2302599  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 34)  
AUTHORS Critchley,I.A., Dodd,I., Barnett,P. and Mossakowska,D.E.  
TITLE D2D3 POLYPEPTIDE FROM STAPHYLOCOCCUS AUREUS AND USES THEREOF  
JOURNAL Patent: WO 9604380-A 8 15-FEB-1996;  
COMMENT SMITHKLINE BEECHAM PLC (GB)  
Other publication AU 3223895 960304.  
source Location/Qualifiers





```

RESULT 13
GORINSORA 90 bp DNA PRI 27-APR-1993
DEFINITION Gorilla endogenous retroviral DNA.
ACCESSION M69096
VERSION M69096.1 GI:177060
KEYWORDS
SOURCE
ORGANISM Gorilla DNA.
Gorilla gorilla
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
1 (bases 1 to 90)
Shih,A., Coutavas,E.E. and Rush,M.G.
Evolutionary implications of primate endogenous retroviruses
JOURNAL Virology 182, 495-502 (1991)
91220699
FEATURES
source
1..90
/organism="Gorilla gorilla"
/db_xref="taxon:9593"
1..90
/partial
/feature="ORF"
/codon_start=1
/protein_id="AA35469.1"
/db_xref="GI:177061"
/translation="FKNSPTLFDALHRLDLDFTRHDPDLILQ"
BASE COUNT 19 a 31 c 15 g 25 t
ORIGIN

Query Match 50.7%; Score 14.2; DB 9; Length 90;
Best Local Similarity 84.2%; Pred. No. 1.8e+04;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cttgacatttgccac 19
Db 45 cttgactgatttgcac 63

RESULT 14
DDIDDK 99 bp DNA INV 27-APR-1993
DEFINITION D.discoidium protein kinase 5 gene, partial cds.
ACCESSION M59748
VERSION M59748.1 GI:167725
KEYWORDS
SOURCE
ORGANISM Dictyostelium discoidium (strain AX-3) DNA.
Dictyostelium discoidium
Eukaryota; Dictyostellida; Dictyostelium.
1 (bases 1 to 99)
Haribabu,B. and Dotlin,R.P.
Identification of a protein kinase multigene family of
Dictyostelium discoidium: Molecular cloning and expression of a
cDNA encoding a developmentally regulated protein kinase
Proc. Natl. Acad. Sci. U.S.A. 88, 1115-1119 (1991)
91142122
FEATURES
source
1..99
/organism="Dictyostelium discoidium"
/strain="AX-3"
/db_xref="taxon:44689"
<1..>99
/codon_start=1
/product="protein kinase 5"
/protein_id="AA33190.1"
/db_xref="GI:167726"
/translation="NILLTSGHIVLTPFGISKGLVSDNDRFATFC"
BASE COUNT 30 a 18 c 24 g 27 t
ORIGIN

Query Match 50.7%; Score 14.2; DB 34; Length 99;

```

```

Best Local Similarity 70.4%; Pred. No. 1.8e+04;
Matches 19; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 ttgacatttggcaccacacataggg 28
Db 34 ttgaccgatttggatpaaagcagag 60

```

```

RESULT 15
E14994/c 27 bp DNA PAT 28-JUL-1999
LOCUS PCR primer to mutagenize equine cDNA encoding chorionic
DEFINITION gonadotropin alpha subunit at the 58th Asn to Gln.
ACCESSION E14994
VERSION E14994.1 GI:5709677
KEYWORDS JP 1998036285-A/8.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 27)
AUTHORS Ogawa,T., Shioda,K., Bin,K. and Ikemi,M.
TITLE NEW GONADOTROPIC HORMONE AND ITS PRODUCTION
JOURNAL Patent: JP 1998036285-A 10-FEB-1998;
DENKI KAGAKU KOGYO KK
COMMENT
OS None
OC Artificial sequences.
PN JP 1998036285-A/8
PD 10-FEB-1998
PE 23-JUL-1996 JP 1996193232
PI OGAWA TOMOYA, SHIODA KUNIO, BIN KANSHIYOKU, IKEMI MASAHIRO
A61K38/24,C07K14/59,C12N5/10,C12N15/09,C12P21/02,C12P21/02, PC
C12R1:91);
CC strandness: Single;
CC topology: Linear;
FH Key
FH Key Location/Qualifiers
FT source 1..27
FT Location/Qualifiers
1..27
/organism="Artificial sequences"
/db_xref="taxon:32644"
BASE COUNT 8 a 8 c 5 g 6 t
ORIGIN

Query Match 50.0%; Score 14; DB 5; Length 27;
Best Local Similarity 77.3%; Pred. No. 2.1e+04;
Matches 17; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 tgcacatttggcaccacacat 24
Db 22 tgatctgcttggaccacacat 1

```

Search completed: June 4, 2000, 16:05:34 /  
Job time: 27886 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 13:53:12 ; Search time 4521.53 Seconds  
(without alignments)  
16.136 Million cell updates/sec

Title: us-09-164-714-12  
Perfect score: 18  
Sequence: 1 ggcgacaacacgacctag 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 119968

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: gb\_est1: \*  
21: gb\_est2: \*  
22: gb\_est3: \*  
23: gb\_est4: \*  
24: gb\_est5: \*  
25: gb\_est6: \*  
26: gb\_est7: \*  
27: gb\_est8: \*  
28: gb\_est9: \*  
29: gb\_est10: \*  
30: gb\_est11: \*  
31: gb\_est12: \*  
32: gb\_est13: \*  
33: gb\_est14: \*  
34: gb\_est15: \*  
35: gb\_est16: \*  
36: gb\_est17: \*  
37: gb\_est18: \*  
38: gb\_est19: \*  
39: gb\_est20: \*  
40: gb\_est21: \*  
41: gb\_est22: \*  
42: gb\_est23: \*  
43: gb\_est24: \*  
44: gb\_est25: \*

45: gb\_est26: \*  
46: gb\_est27: \*  
47: gb\_est28: \*  
48: gb\_est29: \*  
49: gb\_est30: \*  
50: gb\_est31: \*  
51: gb\_est32: \*  
52: em\_est20: \*  
53: em\_est21: \*  
54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*  
59: gb\_est33: \*  
60: gb\_est34: \*  
61: gb\_est35: \*  
62: gb\_est36: \*  
63: gb\_est37: \*  
64: gb\_est38: \*  
65: em\_est27: \*  
66: em\_est28: \*  
67: em\_est29: \*  
68: em\_est30: \*  
69: gb\_est39: \*  
70: gb\_est40: \*  
71: gb\_est41: \*  
72: gb\_est42: \*  
73: gb\_est43: \*  
74: gb\_est44: \*  
75: em\_est31: \*  
76: em\_est32: \*  
77: em\_est33: \*  
78: em\_est34: \*  
79: gb\_est45: \*  
80: gb\_est46: \*  
81: gb\_est47: \*  
82: gb\_gss1: \*  
83: gb\_gss2: \*  
84: gb\_gss3: \*  
85: gb\_gss4: \*  
86: em\_gss1: \*  
87: em\_gss2: \*  
88: em\_gss3: \*  
89: em\_gss4: \*  
90: gb\_gss5: \*  
91: gb\_gss6: \*  
92: gb\_gss7: \*  
93: gb\_gss8: \*  
94: gb\_gss9: \*  
95: em\_gss5: \*  
96: em\_gss6: \*  
97: em\_gss7: \*  
98: em\_gss8: \*  
99: em\_gss9: \*  
100: em\_gss10: \*  
101: em\_gss11: \*  
102: gb\_gss10: \*  
103: gb\_gss11: \*  
104: em\_gss12: \*  
105: gb\_gss12: \*  
106: gb\_gss13: \*  
107: gb\_gss14: \*  
108: gb\_gss15: \*  
109: gb\_gss16: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result

Query

No.	Score	Match	Length	DB	ID	Description
C 1	13.4	74.4	85	25	U44241	U44241 ENU44241 AS
C 2	13.2	73.3	87	40	AA971367	AA971367 op80c11.s
C 3	12.8	71.1	46	36	AA607492	AA607492 vmb1h02.r
C 4	12.8	71.1	97	29	AA151757	AA151757 z029f01.r
C 5	12.8	71.1	97	43	AA1181902	AA1181902 ud73e07.x
C 6	12.4	68.9	53	27	C01203	C01203 HUMGS000790
C 7	12.4	68.9	60	40	AA938614	AA938614 nw91h09.s
C 8	12.4	68.9	67	47	AA1472550	AA1472550 l178a02.x
C 9	12.4	68.9	74	30	AA218675	AA218675 zq14f10.s
C 10	12.4	68.9	86	34	AA515609	AA515609 ng70a03.s
C 11	12.2	67.8	55	43	AA1220035	AA1220035 q978h07.x
C 12	12.2	67.8	79	43	AA1214486	AA1214486 q969d07.x
C 13	12.2	67.8	88	51	AA1744149	AA1744149 wc36h02.x
C 14	12.2	67.8	95	30	AA244496	AA244496 nc07910.s
C 15	12.2	67.8	99	94	AQ025006	AQ025006 EP(2)0969
C 16	12.2	67.8	88	44	AA1310822	AA1310822 q093h04.x
C 17	11.8	65.6	52	40	AA934866	AA934866 op47a10.s
C 18	11.8	65.6	61	39	AA894395	AA894395 of85903.s
C 19	11.8	65.6	66	94	AQ073497	AQ073497 EP(2)2329
C 20	11.8	65.6	74	46	AA1445897	AA1445897 t106905.x
C 21	11.8	65.6	76	38	AA789177	AA789177 aq58c03.s
C 22	11.8	65.6	82	61	AA1865432	AA1865432 wK11h10.x
C 23	11.8	65.6	100	30	AA247764	AA247764 hE0467.s
C 24	11.8	65.6	100	43	AA1204724	AA1204724 ZP-EST40
C 25	11.6	64.4	88	39	AA864661	AA864661 oh37c11.s
C 26	11.6	64.4	91	43	AA181540	AA181540 uc59906.r
C 27	11.6	64.4	94	29	AA144170	AA144170 m954910.r
C 28	11.6	64.4	97	44	AA1323709	AA1323709 mq37c06.x
C 29	11.6	64.4	100	31	AA453587	AA453587 zeh10164
C 30	11.4	63.3	34	38	AA8789803	AA8789803 vL79f06.r
C 31	11.4	63.3	63	44	AA1286217	AA1286217 q102e01.x
C 32	11.4	63.3	65	37	AA733449	AA733449 vL73h08.r
C 33	11.4	63.3	82	34	AA533454	AA533454 nJ68c05.s
C 34	11.2	62.2	37	34	AA466917	AA466917 vL10904.r
C 35	11.2	62.2	37	43	AA1188273	AA1188273 qd11g10.x
C 36	11.2	62.2	44	42	AA1153789	AA1153789 vZ90h06.r
C 37	11.2	62.2	44	94	AQ074027	AQ074027 EP(3)3471
C 38	11.2	62.2	49	28	C20877	C20877 HUMGS000494
C 39	11.2	62.2	50	28	AA108275	AA108275 EST0018.r
C 40	11.2	62.2	51	27	AA000409	AA000409 me76c01.r
C 41	11.2	62.2	51	81	AA424054	AA424054 sh60b02.y
C 42	11.2	62.2	58	34	AA495205	AA495205 fa04c02.r
C 43	11.2	62.2	72	29	AA157222	AA157222 z041d06.r
C 44	11.2	62.2	72	28	AA072178	AA072178 mm69f01.r
C 45	11.2	62.2	73	21	T91290	T91290 yd52c07.s1

## ALIGNMENTS

RESULT 1  
LOCUS U44241 85 bp mRNA EST 03-APR-1996  
DEFINITION ENU44241 Aspergillus nidulans cleistothecium Emericella nidulans  
CDNA clone SE0504, mRNA sequence.  
U44241  
VERSION U44241.1 GI:1244904  
KEYWORDS EST.  
SOURCE Emericella nidulans.  
ORGANISM Eukaryota; Fungi; Ascomycota; Eurotiiales; Trichocomaceae;  
Emericella.

REFERENCE 1 (bases 1 to 85)  
AUTHORS Lee,D., Lee,S., Hwang,H., Kim,J. and Chae,K.  
TITLE Quantitative analysis of gene expression in sexual structures of Aspergillus nidulans by sequencing of 3'-directed cDNA clones  
JOURNAL FEBS Microbiol. Lett. 138 (1), 71-76 (1996)  
MEDLINE 96236220  
COMMENT On Apr 14, 1993 this sequence version replaced gi:838110.  
Contact: Keon-Sang Chae  
Chonbuk National University

Chonju, 561-756, S. Korea  
Tel: +82-652-70-3340  
Fax: +82-652-70-3345  
Email: chae@chonbukns.chonbuk.ac.kr.  
Location/Qualifiers  
1. 85  
/organism="Emericella nidulans"  
/strain="FGSC4"  
/db\_xref="taxon:5072"  
/clone="SE0504"  
/clone\_lib="Aspergillus nidulans cleistothecium"  
/tissue\_type="cleistothecium"  
/cell\_type="Hull cell"  
/dev\_stage="sexual"  
/note="3'-directed cDNA clones; single-pass sequencing"

BASE COUNT 15 a 17 c 22 g 31 t  
ORIGIN

Query Match 74.4%; Score 13.4; DB 25; Length 85;  
Best Local Similarity 93.3%; Pred. No. 3.3e+03;  
Matches 14; conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 acaaacacgactag 18  
|||||  
Db 34 ACAAAACGACCTCG 20

RESULT 2  
LOCUS AA971367 87 bp mRNA EST 07-JUL-1998  
DEFINITION op80c11.s1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
IMAGE:1583156 3', mRNA sequence.  
AA971367  
VERSION AA971367.1 GI:3146657  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 87)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL On Jan 19, 1998 this sequence version replaced gi:2285140.  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert length: 892 Std Error: 0.00  
Seq primer: -40m13 fwd. ER from Amersham  
High quality sequence stop: 73.  
Location/Qualifiers  
1. 87  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1583156"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site:1: Not I; Site:2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung KBHL19W, testis NHT, and B-cell  
NCI-CGAP-GCB1) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."

FEATURES  
source





ACCESSION	A1181902	alpha (mouse);, mRNA sequence.
VERSION	A1181902.1	GI:3732540
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
AUTHORS	Eutheria; Rodentia; Scurionathii; Muridae; Murinae; Mus.	
TITLE	1 (bases 1 to 97)	
JOURNAL	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Waterston,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and	
COMMENT	The WashU-HMI Mouse EST Project Unpublished (1996) On Jan 14, 1998 this sequence version replaced gi:1797360. Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (infoimage.lnl.gov) for further information. MGI:924872	
FEATURES	Trace considered overall poor quality Seq primer: custom primer used High quality sequence stop: 1. location/Qualifiers	
SOURCE	1..97	
	/organism="Mus musculus"	
	/strain="C57BL"	
	/db_xref="taxon:10090"	
	/clone="IMAGE:1451556"	
	/clone_1ib="Sugano mouse liver mlia"	
	/sex="female"	
	/dev_stage="adult"	
	/lab_host="DH10B"	
	/note="Organ: liver; Vector: PME185-FL3; Site:1: DraIII (CACTGTG); Site:2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCCTTTTTCCTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor (TGTTGGGCCACTGTG), digested and cloned into distinct DraIII sites of the pME185-FL3 vector (5' site CACTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTCTGCCTTAAGCTCGG and 3' end primer CGAAGCTGACCTCGACACA."	
BASE COUNT	32 a 19 c 21 g 25 t	
ORIGIN		
Query Match	71.1%; Score 12.8; DB 43; Length 97;	
Best Local Similarity	87.5%; Pred. No. 6.9e+03;	
Matches	14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
OY	1 ggcagcaaacaccagcct 16	
Db	84 GTGACCAAAACCAGCCT 69	
RESULT	6	
C01203/c	C01203 53 bp mRNA EST 23-JUL-1996	
LOCUS	HMMGS0007903 Human adult (K.Okubo) Homo sapiens cDNA, mRNA	
DEFINITION	C01203	
sequence.	C01203.1 GI:1433433	
ACCESSION		
VERSION		

KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE	1 (bases 1 to 53)
JOURNAL	Okubo, K.
COMMENT	BodyMap, human gene expression database Unpublished (1995) Contact: Okubo, K. Institute for Molecular and Cellular Biol Osaka University 1-3,Yamada-oka, Suita, Osaka Pref. 565, Japan Tel.: 06-877-5111(ext.3315) Email: kousaku@imcb.osaka-u.ac.jp Human Gene Signature, 3'-directed cDNA sequence. We are not submitting the same cDNA sequence redundantly to DBD since 1993. For the abundance information of clones with this sequence in this library and as well as in other 3'-directed libraries, see ' <a href="http://www.imcb.osaka-u.ac.jp/bodymap/">http://www.imcb.osaka-u.ac.jp/bodymap/</a> '. The sequences of the clones represented by this GS sequences is also found there.
FEATURES	Location/Qualifiers 1..53 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="Human adult (K.Okubo)" /dev_stage="adult"
BASE COUNT	10 a                15 c                16 g                12 t
ORIGIN	
Query Match	68.9%; Score 12.4; DB 27; Length 53;
Best Local Similarity	92.9%; Pred. No. 1e+04;
Matches 13; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
Oy	5 caaacacgacctag 18 
Db	32 CAACAACGACCAG 19
RESULT 7	
LOCUS	AA938614 60 bp mRNA EST 30-APR-1998
DEFINITION	nw11h09.s1 NCI-CGAP_Pri12 Homo sapiens CDNA clone IMAGE:1254017
ACCESSION	AA938614 similar to gb:U011120 GLUCOSE-6-PHOSPHATASE (HUMAN); mRNA sequence.
VERSION	AA938614.1 GI:3096642
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE	1 (bases 1 to 60)
JOURNAL	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) On Jan 19, 1998 this sequence version replaced gi:2152774. Contact: Robert Strausberg, Ph.D. Tel.: (301) 496-1550 Email: Robert_Strausberg@nih.gov Tissue Procurement: W. Douglas Fligg, Ph.D., Paul H. Duray, M.D., Rodrigo F. Chuquai, M.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: David B. Krizman, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/JLML at: <a href="http://www-bio.llnl.gov/bbrp/image/image.html">www-bio.llnl.gov/bbrp/image/image.html</a>

FEATURES  
source  
Location/Qualifiers  
1. .60  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1254017"  
/clone\_lib="NCI\_CGAP\_Pri12"  
/sex="male"  
/tissue\_type="metastatic prostate bone lesion"  
/lab\_host="DH10B"  
/note="Vector: PAMPI0; mRNA made from metastatic prostate lesion of the bone, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Library made by D. Krizman, NIH."

BASE COUNT  
ORIGIN  
9 a 13 c 13 g 25 t

Query Match  
Best Local Similarity 92.9%; Score 12.4; DB 40; Length 60;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 caaacacgacctag 18  
|||||

Db 48 CAAGACCAAGCCTAG 35

RESULT 8  
AI472550/c  
LOCUS t178a02.x1 NCI\_CGAP\_Col4 Homo sapiens cDNA clone IMAGE:2153162 3',  
DEFINITION mRNA sequence.  
ACCESSION AI472550  
VERSION AI472550.1 GI:4325595  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 67)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jun 5, 1998 this sequence version replaced gi:3187297.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/dbp/submit/submit.html

FEATURES  
source  
Insert Length: 1410 Std Error: 0.00  
Seg primer: -40UP from gibco  
High quality sequence stop: 61.  
Location/Qualifiers  
1. .67  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2153162"  
/clone\_lib="NCI\_CGAP\_Col4"  
/tissue\_type="moderately-differentiated adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: PCMV-SPORE6; Site: 1: SalI;  
Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.7 kb. Life Technologies catalog #: 11531-019"

BASE COUNT  
ORIGIN  
11 a 16 c 15 g 25 t

ORIGIN  
Query Match  
Best Local Similarity 92.9%; Score 12.4; DB 47; Length 67;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 caaacacgacctag 18  
|||||

Db 41 CAAGACCAAGCCTAG 28

RESULT 9  
AA218675/c  
LOCUS zq14f10.s1 Stratagene fetal retina 937202 Homo sapiens cDNA clone  
DEFINITION IMAGE:629707 3' similar to contains Alu repetitive element; mRNA  
sequence.  
ACCESSION AA218675  
VERSION AA218675.1 GI:1832759  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 74)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapell, B.,  
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rife, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierly-Meg, J.,  
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,  
and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
TITLE  
JOURNAL  
MEDLINE  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1393955.  
Contact: Wilson R.  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seg primer: -41m3 fwd. ET from Amersham  
High quality sequence stop: 51.  
Location/Qualifiers  
1. .74  
/organism="Homo sapiens"  
/db\_xref="GDB:517979"  
/db\_xref="taxon:9606"  
/clone="IMAGE:629707"  
/clone\_lib="Stratagene fetal retina 937202"  
/sex="mixed"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: pBluescript SK-; Site: 1: EcoRI; Site: 2:  
XhoI; Cloned unidirectionally. Primer: Oligo dT. Pooled  
retinal tissue. Average insert size: 1.0 kb; Uni-ZAP XR  
Vector: -5' adaptor sequence: 5' GAAATCGGACGAG 3' -3'  
adaptor sequence: 5' CTCGAGCTTTT TTTT TTTT TTTT 3"

BASE COUNT  
ORIGIN  
15 a 18 c 16 g 24 t

Query Match  
Best Local Similarity 92.9%; Score 12.4; DB 30; Length 74;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 caaacacgacctag 18  
|||||

Db 46 CAAGACCAAGCCTAG 33

RESULT 10  
AA515609 86 bp mRNA EST 19-AUG-1997  
LOCUS NCI-CGAP\_Lip2 Homo sapiens cDNA clone IMAGE:940108  
DEFINITION similar to contains Alu repetitive element, contains element THR  
repetitive element ; mRNA sequence.  
ACCESSION AA515609  
VERSION AA515609.1 GI:2255209  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 86)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On May 9, 1995 this sequence version replaced gi:802649.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
[www-bio.lnl.gov/dbp/image/image.html](http://www-bio.lnl.gov/dbp/image/image.html)

FEATURES  
SOURCE  
Insert Length: 591 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 76.  
Location/Qualifiers  
1..86  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:940108"  
/clone\_lib="NCI-CGAP\_Lip2"  
/tissue\_type="liposarcoma"  
/lab\_host="DH10B"  
/note="Vector: PAMPI0; mRNA made from liposarcoma, CDNA  
made by oligo-dT priming. Non-directionally cloned.  
Size-selected on agarose gel, average insert size 600 bp.  
Reference: Krizman et al. (1996) Cancer Research  
56:5380-5383."

BASE COUNT 28 a 21 c 26 g 11 t  
ORIGIN

Query Match 68.9%; Score 12.4; DB 34; Length 86;  
Best Local Similarity 92.9%; Pred. No. 1.1e+04;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 caaacacgacctag 18  
|||||  
Db 41 CAAACGACGCTCG 54

RESULT 11  
AI220035 55 bp mRNA EST 30-NOV-1998  
LOCUS qg78h07.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
DEFINITION IMAGE:1841341 3' similar to SM:ZNI6\_HUMAN P17020 ZINC FINGER  
PROTEIN 16 ; mRNA sequence.  
ACCESSION AI220035  
VERSION AI220035.1 GI:3802238  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE 1 (bases 1 to 55)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([infoimage.lnl.gov](http://infoimage.lnl.gov)) for further information.  
Insert Length: 1198 Std Error: 0.00  
Seq primer: -40UP from Gldco  
High quality sequence stop: 1.  
Location/Qualifiers  
1..55  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1841341"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT7n3D-Pac (Pharmacia) with  
a modified polylinker; Site:1: Not I; Site:2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBHL19W, testis NHT, and B-cell  
NCI-CGAP.CCB1) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries.  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bernaldo."

BASE COUNT 10 a 21 c 17 g 7 t  
ORIGIN

Query Match 67.8%; Score 12.2; DB 43; Length 55;  
Best Local Similarity 82.4%; Pred. No. 1.3e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggcacaaacgacctc 17  
|||||  
Db 30 GCTGCAAAACCAACCTA 46

RESULT 12  
AI214486 79 bp mRNA EST 21-OCT-1998  
LOCUS q959d07.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
DEFINITION IMAGE:1840429 3' similar to TR:O14978 O14978 ZINC FINGER PROTEIN  
FPM315 ; mRNA sequence.  
ACCESSION AI214486  
VERSION AI214486.1 GI:3778087  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 79)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 17, 1998 this sequence version replaced gi:2043831.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([infoimage.lnl.gov](http://infoimage.lnl.gov)) for further information.  
Seq primer: -40UP from Gldco

## FEATURES High quality sequence stop: 1.

Location/Qualifiers

1..79

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1840429"

/clone.lib="Soares\_NFL\_T\_GBC\_S1"

/lab\_host="DH10B"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbH19w, testis NHT, and B-cell NCI CGAP GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 16 a 33 c 15 g 15 t

ORIGIN

Query Match 67.8%; Score 12.2; DB 43; Length 79;  
Best Local Similarity 82.4%; Pred. No. 1.4e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ggcagaaaccagccta 17  
|| ||||| ||||  
Db 30 GCCTCAAMCACCACCTA 46

RESULT 13  
AI744149 88 bp mRNA EST 17-DEC-1999  
LOCUS WC36b02.x1 NCI\_CGAP\_Pr28 Homo sapiens cDNA clone IMAGE:2317299 3,  
DEFINITION similar to TR:Q13765 Q13765 NASCENT POLYPEPTIDE ASSOCIATED COMPLEX  
ALPHA SUBUNIT.; mRNA sequence.

ACCESSION AI744149  
VERSION AI744149.1 GI:5112437  
KEYWORDS EST.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 88)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>

Tumor Gene Index  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL

Unpublished (1997)  
On Dec 20, 1995 this sequence version replaced gi:1134761.

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

[www-bio.llnl.gov/dbfp/image/image.html](http://www-bio.llnl.gov/dbfp/image/image.html)

Trace considered overall poor quality

Insert Length: 829 Std Error: 0.00

Seq primer: 400P from Gibco

High quality sequence stop: 1.

FEATURES

source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2317299"

/clone.lib="NCI\_CGAP\_Pr28"

/sex="male"

/dev\_stage="adult"

/lab\_host="DH10B"

/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI CGAP Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonids 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 9 a 24 c 20 g 34 t 1 others

ORIGIN

Query Match 67.8%; Score 12.2; DB 51; Length 88;  
Best Local Similarity 82.4%; Pred. No. 1.4e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 cgacaaaccagcctag 18  
||| ||||| ||  
Db 84 CCACACACACGCCCCAG 68

RESULT 14  
AA244496 95 bp mRNA EST 20-AUG-1997  
LOCUS nc07g10.s1 NCI\_CGAP\_Pr1 Homo sapiens cDNA clone IMAGE:1007490, mRNA  
DEFINITION sequence.

ACCESSION AA244496  
VERSION AA244496.1 GI:1875215  
KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 95)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>

Tumor Gene Index  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Unpublished (1997)

On Apr 14, 1993 this sequence version replaced gi:503300.

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuquiqui, M.D., Michael Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: David B. Kitzman, Ph.D.

DNA Sequencing by: Genome Systems Inc., Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

[www-bio.llnl.gov/dbfp/image/image.html](http://www-bio.llnl.gov/dbfp/image/image.html)

Seq primer: -41m13 fwd. ET from Amersham

High quality sequence stop: 82.

Location/Qualifiers

1..95

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1007490"

/clone.lib="NCI\_CGAP\_Pr1"

/sex="Male"

/dev\_stage="45 years old"

/lab\_host="DH10B"

/note="Vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st strand cDNA was primed with oligo(dT)17 on 50 ng of DNase-treated, total cellular RNA obtained from 5,000-10,000 microdissected, histologically normal prostate epithelial cells. Double-stranded cDNA was ligated to EcoRI adaptors, 5 cycles of PCR applied to the

cdna with an adaptor-specific primer, and the resulting PCR product subcloned into pAMP10 by the UDC-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman."

BASE COUNT 28 a 13 c 27 g 27 t

ORIGIN

Query Match 67.8%; Score 12.2; DB 30; Length 95;  
Best Local Similarity 82.4%; Pred. No. 1.4e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 gcagacaacacagccta 17  
|||  
Db 71 GCACCAAAACCACTCTA 55

## RESULT 15

A0025006

LOCUS

DEFINITION

A0025006 99 bp DNA GSS 14-OCT-1998  
EP(2)0969 Drosophila melanogaster EP line Drosophila melanogaster  
genomic Sequence recovered from 5' end of P element, genomic survey  
sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

The BOSP gene disruption project: single EP element insertions  
unpublished (1998)  
Contact: Gerald Rubin  
Berkeley Drosophila Genome Project  
University of California, Berkeley  
LSA Building, Berkeley, CA 94720-3200, USA  
Fax: 5106439947  
Email: gerry@fruitfly.berkeley.edu

Sequence recovery method was inverse PCR.

Sequence orientation is forward strand relative to 5' end of P  
element

The P element insertion position is base 92 in the 99 bases. This  
insertion position refers to the first base of the 8 base target  
recognition sequence.

Class: transposon-tagged.

FEATURES

SOURCE

Location/Qualifiers

1..99

/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"

/clone.lib="Drosophila melanogaster EP line"

/note="Inverse PCR was performed on Drosophila  
melanogaster strains each of which contains a single EP  
transposable element insertion. (The generation of these  
insertion strains is described in North P, Szabo K, Bailey  
A, Laverly T, Rehm J, Rubin GW, Weigmann K, Milan M, Benes  
V, Ansoorge W, Cohen SM. 1998. Systematic gain-of-function  
genetics in Drosophila. Development 6:1049-1057.) The  
resultant fragment for each strain was directly sequenced  
to determine the genomic sequence at the site of  
insertion. Details of the protocols used can be found at  
[http://fruitfly.berkeley.edu/p-disrupt/inverse\\_pcr.html](http://fruitfly.berkeley.edu/p-disrupt/inverse_pcr.html)."

23 a 23 c 20 g 33 t

BASE COUNT  
ORIGIN

Query Match 67.8%; Score 12.2; DB 94; Length 99;  
Best Local Similarity 82.4%; Pred. No. 1.4e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 cgacaacacagcctag 18  
|||  
Db 37 CAACAAACCAAGGAG 53

Search completed: June 4, 2000, 13:53:16  
Job time: 20940 sec

---

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:05:34 ; Search time 1236.38 Seconds  
(without alignments)  
-14.163 Million cell updates/sec

Title: US-09-164-714-12  
Perfect score: 18  
Sequence: 1 gcgcacaaccagcctag 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 356616

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenBml :  
1: gb\_ba1 :  
2: gb\_ba2 :  
3: gb\_om :  
4: gb\_om :  
5: gb\_om :  
6: gb\_ph :  
7: gb\_pl1 :  
8: gb\_pl2 :  
9: gb\_pl1 :  
10: gb\_pl2 :  
11: gb\_pl3 :  
12: gb\_ro :  
13: gb\_sts :  
14: gb\_sy :  
15: gb\_un :  
16: gb\_vl :  
17: em\_fun :  
18: em\_hum1 :  
19: em\_hum2 :  
20: em\_in :  
21: em\_om :  
22: em\_or :  
23: em\_ov :  
24: em\_pat :  
25: em\_ph :  
26: em\_pl :  
27: em\_ro :  
28: em\_sts :  
29: em\_sy :  
30: em\_un :  
31: em\_vl :  
32: gb\_hcg1 :  
33: gb\_hcg2 :  
34: gb\_hcg3 :  
35: gb\_hcg4 :  
36: em\_ba1 :  
37: em\_ba2 :  
38: em\_hum3 :  
39: em\_hum4 :  
40: gb\_pr4 :  
41: gb\_hcg3 :  
42: gb\_hcg4 :  
43: gb\_hcg5 :  
44: gb\_hcg6 :

45: gb\_hcg7 :  
46: em\_hcg1 :  
47: em\_hcg2 :  
48: em\_hcg3 :  
49: em\_hum5 :  
50: gb\_pl3 :  
51: gb\_pr5 :  
52: gb\_hcg8 :  
53: gb\_hcg9 :  
54: gb\_hcg10 :  
55: gb\_hcg11 :  
56: gb\_hcg12 :  
57: gb\_hcg13 :  
58: gb\_hcg14 :  
  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	13.8	76.7	25	5	E07840	E07840 Upstream ac
C 2	13.8	76.7	25	5	I72301	I72301 Sequence 10
C 3	13.8	76.7	25	5	I79557	I79557 Sequence 2
C 4	13.8	76.7	30	5	I72318	I72318 Sequence 27
C 5	13.8	76.7	31	5	I72311	I72311 Sequence 20
C 6	13.8	76.7	31	5	I72312	I72312 Sequence 21
C 7	13.8	76.7	31	5	I79569	I79569 Sequence 14
C 8	13.8	76.7	31	5	I79570	I79570 Sequence 15
C 9	13.8	76.7	50	5	I72317	I72317 Sequence 26
C 10	13.4	74.4	30	5	A36554	A36554 Sequence 8
C 11	13.2	73.3	78	12	M0932	M0932 Mouse DNA u
C 12	12.4	68.9	30	5	A36555	A36555 Sequence 9
C 13	12.2	67.8	36	5	A19278	A19278 oligonucleo
C 14	12.2	67.8	39	5	A36995	A36995 Sequence 46
C 15	12.2	67.8	39	5	A39203	A39203 Sequence 51
C 16	12.2	67.8	39	5	A39262	A39262 Sequence 46
C 17	12.2	67.8	49	5	A16464	A16464 oligonucleo
C 18	12.2	67.8	49	5	A16465	A16465 oligonucleo
C 19	12.2	67.8	50	5	I79573	I79573 Sequence 18
C 20	12.2	67.8	50	5	I79574	I79574 Sequence 19
C 21	12.2	67.8	55	5	A19300	A19300 oligonucleo
C 22	12.2	67.8	57	5	E12104	E12104 signal pept
C 23	12.2	67.8	57	5	I04131	I04131 Sequence 19
C 24	12.2	67.8	57	5	I05343	I05343 Sequence 1
C 25	12.2	67.8	57	5	I06744	I06744 Sequence 12
C 26	12.2	67.8	63	7	YSC50C2B	M13627 Yeast (S.ce
C 27	12.2	67.8	64	5	A23341	A23341 Artificial
C 28	12.2	67.8	64	5	AR068519	AR068519 Sequence
C 29	12.2	67.8	68	14	SYNVSCTNV	M13346 Yeast (S.ce
C 30	12.2	67.8	69	5	AR038140	AR038140 Sequence
C 31	12.2	67.8	69	5	I79868	I79868 Sequence 8
C 32	12.2	67.8	70	5	A24425	A24425 SCFVla link
C 33	12.2	67.8	75	5	E00281	E00281 DNA coding
C 34	12.2	67.8	75	5	E00283	E00283 DNA coding
C 35	12.2	67.8	84	5	A42985	A42985 Sequence 20
C 36	12.2	67.8	84	5	A45530	A45530 Sequence 15
C 37	12.2	67.8	84	5	AR030802	AR030802 Sequence
C 38	12.2	67.8	94	5	I33294	I33294 Sequence 26
C 39	12	66.7	77	11	AF032263	AF032263 Lemur sp.
C 40	11.8	65.6	21	5	A60234	A60234 Sequence 1
C 41	11.8	65.6	30	5	A36556	A36556 Sequence 10
C 42	11.8	65.6	30	5	A36557	A36557 Sequence 11
C 43	11.8	65.6	30	5	A36558	A36558 Sequence 12
C 44	11.8	65.6	30	5	A60242	A60242 Sequence 9
C 45	11.8	65.6	38	5	A11228	A11228 Oligonucleo

## ALIGNMENTS

RESULT 1  
E07840/c 25 bp DNA PAT 29-SEP-1997  
LOCUS Upstream activation sequence in AOX2 promoter(Alcohol Oxidase 2  
DEFINITION promoter).  
ACCESSION E07840  
VERSION E07840.1 GI:2175973  
KEYWORDS JP 1994189769-A/10.  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 25)  
AUTHORS Oh,H., Miura,M., Uno,S., Chuganji,M., Hiramatsu,T., and Omura,T.  
TITLE VARIANT AOX2 PROMOTER, VECTOR CARRYING THE PROMOTER, TRANSFORMANT  
JOURNAL AND PRODUCTION OF FOREIGN PROTEIN  
PATENT: JP 1994189769-A 10 12-JUL-1994;  
GREEN CROSS CORP:THE  
COMMENT OS None  
OC Artificial sequences.  
PN JP 1994189769-A/10  
PD 12-JUL-1994  
PE 06-AUG-1993 JP 1993215306  
PR 30-OCT-1992 JP 92P 293315  
PI OI HIDEYUKI, MIURA MASAMI, UNO SHUSEI, CHUGENJI MASAKO, PI  
HIRAMATSU TAKAO  
PI OMURA TAKAO  
PC C12N15/14,C12N1/19,C12N15/67,C12N15/81,C12P21/02,(C12N1/19, PC  
C12R1:84)  
PC (C12P21/02,C12R1:84);  
CC strandedness: Double;  
CC topology: Linear;  
FH Key Location/Qualifiers  
FT source 1..25  
FT Location/Qualifiers  
FEATURES  
source 1..25  
/organism="Artificial sequences"  
/db\_xref="taxon:32644"  
BASE COUNT 7 a 3 c 5 g 10 t  
ORIGIN

Query Match 76.7%; Score 13.8; DB 5; Length 25;  
Best Local Similarity 88.2%; Pred. No. 1.1e+03;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gcgacaaacagccta 17  
|||||  
Db 19 GCGACAAAATAGCCTA 3

RESULT 2  
I72301/c 25 bp DNA PAT 03-APR-1998  
LOCUS I72301  
DEFINITION Sequence 10 from patent US 5683893.  
ACCESSION I72301  
VERSION I72301.1 GI:3008440  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 25).  
AUTHORS Oh,H., Miura,M., Uno,S., Chuganji,M., Hiramatsu,R. and Ohmura,T.  
TITLE Mutant AOX2 promoter, vector carrying same, transformant, and  
JOURNAL Production of heterologous Protein  
PATENT: US 5683893-A 10 04-NOV-1997;  
FEATURES Location/Qualifiers  
source 1..25  
/organism="unknown"  
BASE COUNT 7 a 3 c 5 g 10 t  
ORIGIN

Query Match 76.7%; Score 13.8; DB 5; Length 25;  
Best Local Similarity 88.2%; Pred. No. 1.1e+03;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gcgacaaacagccta 17  
|||||  
Db 19 GCGACAAAATAGCCTA 3

RESULT 3  
I79557/c 25 bp DNA PAT 10-JUN-1998  
LOCUS I79557  
DEFINITION Sequence 2 from patent US 5707827.  
ACCESSION I79557  
VERSION I79557.1 GI:3207847  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 25)  
AUTHORS Oh,H., Miura,M., Hiramatsu,R. and Ohmura,T.  
TITLE Mutant AOX2 promoter, vector carrying same, transformant and  
JOURNAL Production of heterologous protein  
PATENT: US 5707827-A 2 13-JAN-1998;  
FEATURES Location/Qualifiers  
source 1..25  
/organism="unknown"  
BASE COUNT 7 a 3 c 5 g 10 t  
ORIGIN

Query Match 76.7%; Score 13.8; DB 5; Length 25;  
Best Local Similarity 88.2%; Pred. No. 1.1e+03;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gcgacaaacagccta 17  
|||||  
Db 19 GCGACAAAATAGCCTA 3

RESULT 4  
I72318/c 30 bp DNA PAT 03-APR-1998  
LOCUS I72318  
DEFINITION Sequence 27 from patent US 5683893.  
ACCESSION I72318  
VERSION I72318.1 GI:3008457  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 30)  
AUTHORS Oh,H., Miura,M., Uno,S., Chuganji,M., Hiramatsu,R. and Ohmura,T.  
TITLE Mutant AOX2 promoter, vector carrying same, transformant, and  
JOURNAL Production of heterologous protein  
PATENT: US 5683893-A 27 04-NOV-1997;  
FEATURES Location/Qualifiers  
source 1..30  
/organism="unknown"  
BASE COUNT 9 a 5 c 5 g 11 t  
ORIGIN

Query Match 76.7%; Score 13.8; DB 5; Length 30;  
Best Local Similarity 88.2%; Pred. No. 1.1e+03;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gcgacaaacagccta 17  
|||||  
Db 23 GCGACAAAATAGCCTA 7

RESULT 5  
LOCUS 172311/c 31 bp DNA PAT 03-APR-1998  
DEFINITION Sequence 20 from patent US 5683893.  
ACCESSION 172311  
VERSION 172311.1 GI:3008450  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 31)  
AUTHORS Ohl,H., Miura,M., Uno,S., Chuganji,M., Hiramatsu,R. and Ohmura,T.  
TITLE Mutant AOX2 promoter, vector carrying same, transformant, and production of heterologous protein  
JOURNAL Patent: US 5683893-A 20 04-NOV-1997;  
FEATURES Location/Qualifiers  
source 1..31  
/organism="unknown"  
BASE COUNT 9 a 4 c 6 g 12 t  
ORIGIN

Query Match 76.7%; Score 13.8; DB 5; Length 31;  
Best Local Similarity 88.2%; Pred. No. 1.1e+03;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gcgacaaacccagccta 17  
|||||  
Db 24 GCGACAAAATAGCCTA 8

RESULT 6  
LOCUS 172312 31 bp DNA PAT 03-APR-1998  
DEFINITION Sequence 21 from patent US 5683893.  
ACCESSION 172312  
VERSION 172312.1 GI:3008451  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 31)  
AUTHORS Ohl,H., Miura,M., Uno,S., Chuganji,M., Hiramatsu,R. and Ohmura,T.  
TITLE Mutant AOX2 promoter, vector carrying same, transformant, and production of heterologous protein  
JOURNAL Patent: US 5683893-A 21 04-NOV-1997;  
FEATURES Location/Qualifiers  
source 1..31  
/organism="unknown"  
BASE COUNT 12 a 6 c 4 g 9 t  
ORIGIN

Query Match 76.7%; Score 13.8; DB 5; Length 31;  
Best Local Similarity 88.2%; Pred. No. 1.1e+03;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gcgacaaacccagccta 17  
|||||  
Db 12 GCGACAAAATAGCCTA 28

RESULT 7  
LOCUS 179569/c 31 bp DNA PAT 10-JUN-1998  
DEFINITION Sequence 14 from patent US 5707827.  
ACCESSION 179569  
VERSION 179569.1 GI:3207859  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 31)  
AUTHORS Ohl,H., Miura,M., Uno,S., Chuganji,M., Hiramatsu,R. and Ohmura,T.  
TITLE Mutant AOX2 promoter, vector carrying same, transformant, and production of heterologous protein  
JOURNAL Patent: US 5683893-A 26 04-NOV-1997;  
FEATURES Location/Qualifiers  
source 1..50  
/organism="unknown"  
BASE COUNT 15 a 9 c 7 g 19 t  
ORIGIN

AUTHORS Ohl,H., Miura,M., Hiramatsu,R. and Ohmura,T.  
TITLE Mutant AOX2 promoter, vector carrying same, transformant and production of heterologous protein  
JOURNAL Patent: US 5707827-A 14 13-JAN-1998;  
FEATURES Location/Qualifiers  
source 1..31  
/organism="unknown"  
BASE COUNT 9 a 4 c 6 g 12 t  
ORIGIN

Query Match 76.7%; Score 13.8; DB 5; Length 31;  
Best Local Similarity 88.2%; Pred. No. 1.1e+03;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gcgacaaacccagccta 17  
|||||  
Db 24 GCGACAAAATAGCCTA 8

RESULT 8  
LOCUS 179570 31 bp DNA PAT 10-JUN-1998  
DEFINITION Sequence 15 from patent US 5707827.  
ACCESSION 179570  
VERSION 179570.1 GI:3207860  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 31)  
AUTHORS Ohl,H., Miura,M., Hiramatsu,R. and Ohmura,T.  
TITLE Mutant AOX2 promoter, vector carrying same, transformant and production of heterologous protein  
JOURNAL Patent: US 5707827-A 15 13-JAN-1998;  
FEATURES Location/Qualifiers  
source 1..31  
/organism="unknown"  
BASE COUNT 12 a 6 c 4 g 9 t  
ORIGIN

Query Match 76.7%; Score 13.8; DB 5; Length 31;  
Best Local Similarity 88.2%; Pred. No. 1.1e+03;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gcgacaaacccagccta 17  
|||||  
Db 12 GCGACAAAATAGCCTA 28

RESULT 9  
LOCUS 172317/c 50 bp DNA PAT 03-APR-1998  
DEFINITION Sequence 26 from patent US 5683893.  
ACCESSION 172317  
VERSION 172317.1 GI:3008456  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 50)  
AUTHORS Ohl,H., Miura,M., Uno,S., Chuganji,M., Hiramatsu,R. and Ohmura,T.  
TITLE Mutant AOX2 promoter, vector carrying same, transformant, and production of heterologous protein  
JOURNAL Patent: US 5683893-A 26 04-NOV-1997;  
FEATURES Location/Qualifiers  
source 1..50  
/organism="unknown"  
BASE COUNT 15 a 9 c 7 g 19 t  
ORIGIN

Query Match 76.7%; Score 13.8; DB 5; Length 50;  
Best Local Similarity 88.2%; Pred. No. 1.2e+03;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gcgacaaaccagccta 17  
|||||  
Db 41 GCGACAAAATAGCCTA 25

RESULT 10  
A36554/c 30 bp DNA PAT 05-MAR-1997

LOCUS A36554  
DEFINITION Sequence 8 from Patent W09325708.  
ACCESSION A36554  
VERSION A36554.1 GI:2293865

KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
1 (bases 1 to 30)  
Darrasse, A., Kocoujansky, A. and Bertheau, Y.  
NUCLEOTIDIC SEQUENCES OBTAINED FROM GENES CODING PEPTATASE-LYASES,  
AND UTILIZATIONS THEREOF PARTICULARLY FOR THE DETECTION OF BACTERIA  
OF THE GENUS -*(ERWINIA)*  
Patent: WO 9325708-A 8 23-DEC-1993;  
AGRONOMIQUE INST NAT RECH (FR)  
Other publication CA 2137253 931223  
Other publication FR 2691978 931210.

JOURNAL  
COMMENT  
FEATURES  
1. .30  
Location/Qualifiers  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 3 a 6 c 11 g 10 t  
ORIGIN

Query Match 74.4%; Score 13.4; DB 5; Length 30;  
Best Local Similarity 93.3%; Pred. No. 1.9e+03;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 cgacaaccagcct 16  
|||||  
Db 27 CGACAAACCATCT 13

RESULT 11  
MUSENICO 78 bp DNA ROD 27-APR-1993

LOCUS  
DEFINITION Mouse DNA upstream of Ig lambda-2 with an insertion site for  
transposon Etn.  
ACCESSION M30932  
VERSION M30932.1 GI:193188

KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
1 (bases 1 to 78)  
Weiss, S. and Johansson, B.  
Integration of the transposon-like element Etn upstream of  
V-lambda-2 in the cell line P3X63Ag8

JOURNAL  
MEDLINE  
FEATURES  
1. .78  
Location/Qualifiers  
/organism="Mus musculus"  
/db\_xref="taxon:10090"

BASE COUNT 33 a 13 c 15 g 17 t  
ORIGIN

Query Match 73.3%; Score 13.2; DB 12; Length 78;  
Best Local Similarity 83.3%; Pred. No. 2.8e+03;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gcgacaaaccagcctag 18  
|||||  
Db 22 GCAACAAAACACCTAG 39

RESULT 12  
A36555/c 30 bp DNA PAT 05-MAR-1997

LOCUS A36555  
DEFINITION Sequence 9 from Patent W09325708.  
ACCESSION A36555  
VERSION A36555.1 GI:2293866

KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
1 (bases 1 to 30)  
Darrasse, A., Kocoujansky, A. and Bertheau, Y.  
NUCLEOTIDIC SEQUENCES OBTAINED FROM GENES CODING PEPTATASE-LYASES,  
AND UTILIZATIONS THEREOF PARTICULARLY FOR THE DETECTION OF BACTERIA  
OF THE GENUS -*(ERWINIA)*  
Patent: WO 9325708-A 9 23-DEC-1993;  
AGRONOMIQUE INST NAT RECH (FR)  
Other publication CA 2137253 931223  
Other publication FR 2691978 931210.

JOURNAL  
COMMENT  
FEATURES  
1. .30  
Location/Qualifiers  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 5 a 6 c 9 g 10 t  
ORIGIN

Query Match 68.9%; Score 12.4; DB 5; Length 30;  
Best Local Similarity 92.9%; Pred. No. 7.5e+03;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 gacaaccagcct 16  
|||||  
Db 26 GACAAACCATCT 13

RESULT 13  
A19278/c 36 bp DNA PAT 01-JUN-1994

LOCUS A19278  
DEFINITION Oligonucleotide.  
ACCESSION A19278  
VERSION A19278.1 GI:513506

KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
1 (bases 1 to 36)  
XYLANASE PRODUCTION  
Patent: WO 9119782-A 18 26-DEC-1991;

JOURNAL  
FEATURES  
1. .36  
Location/Qualifiers  
/organism="synthetic construct"  
/db\_xref="taxon:32630"

BASE COUNT 6 a 8 c 6 g 16 t  
ORIGIN

Query Match 67.8%; Score 12.2; DB 5; Length 36;  
Best Local Similarity 82.4%; Pred. No. 1e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gcgacaaaccagccta 17  
|||||

Db 29 GCTGCAAAACGACCA 13

RESULT 14

A36995/c A36995 39 bp DNA PAT 05-MAR-1997

DEFINITION Sequence 46 from Patent WO9403578.

ACCESSION A36995

VERSION A36995.1 GI:2294189

KEYWORDS

SOURCE unidentified.

ORGANISM unidentified.

REFERENCE 1 (bases 1 to 39)

AUTHORS Hildgen,H.T., Marugy,J.D., Warr,J.F., Klugkist,J., Musters,W. and Hordmann,D.H.

TITLE ENZYMATIC DETERGENT COMPOSITIONS

JOURNAL Patent: WO 9403578-A 46 17-FEB-1994;

COMMENT UNILEVER PLC (GB)

Other publication CZ 9500242 951018

Other publication CN 1088256 940622

Other publication PL 307269 950515

Other publication CA 2141559 940217

Other publication AU 4700793 940303

Other publication ZA 9305530 950130

Other publication JP 8502084T 960305.

FEATURES

source 1..39

/organism="unidentified"

/db\_xref="taxon:32644"

BASE COUNT 8 a 7 c 11 g 13 t

ORIGIN

Query Match 67.8%; Score 12.2; DB 5; Length 39;

Best Local Similarity 82.4%; Pred. No. 1e+04; 3; Indels 0; Gaps 0;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gcgacaaaccagccta 17

|| |||||

Db 20 GCTGCAAAACGACCA 4

RESULT 15

A39203/c A39203 39 bp DNA PAT 05-MAR-1997

DEFINITION Sequence 51 from Patent WO9414963.

ACCESSION A39203

VERSION A39203.1 GI:2295572

KEYWORDS

SOURCE unidentified.

ORGANISM unidentified.

REFERENCE 1 (bases 1 to 39)

AUTHORS Edmond,M.R., Van,D.H., Musters,W., Peters,H., Verrips,C.T. and De,V.J.

TITLE MODIFIED CUTINASES, DNA, VECTOR AND HOST

JOURNAL Patent: WO 9414963-A 51 07-JUL-1994;

COMMENT UNILEVER PLC (GB)

Other publication SK 79595 951108

Other publication PL 309388 951002

Other publication CA 2150837 940707

Other publication AU 5699994 940719

Other publication CN 1090328 940803

Other publication CZ 9501578 951213

Other publication ZA 9309415 950615.

FEATURES

source 1..39

/organism="unidentified"

/db\_xref="taxon:32644"

BASE COUNT 8 a 7 c 11 g 13 t

ORIGIN

Query Match 67.8%; Score 12.2; DB 5; Length 39;

Best Local Similarity 82.4%; Pred. No. 1e+04; 3; Indels 0; Gaps 0;

Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gcgacaaaccagccta 17

|| |||||

Db 20 GCTGCAAAACGACCA 4

Search completed: June 4, 2000, 16:05:35

Job time: 27887 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:09:42 ; Search time 244.64 Seconds  
(without alignments)  
14.877 Million cell updates/sec

Title: US-09-164-714-11

Perfect score: 28

Sequence: 1 ctggaccaatttggcaccacacatagg 28

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 375880

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/PCBUS.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/ina/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	57.1	39	1	US-08-361-920-32 Sequence 32, Appl
2	16	57.1	39	1	US-08-479-939-32 Sequence 32, Appl
3	16	57.1	39	2	US-08-483-433-32 Sequence 32, Appl
4	15	53.6	34	4	US-08-459-135A-2 Sequence 2, Appl
5	14.2	50.7	72	1	US-08-198-431-31 Sequence 31, Appl
6	14.2	50.7	72	2	US-08-564-955-31 Sequence 31, Appl
7	14.2	50.7	72	2	US-08-537-874-31 Sequence 31, Appl
8	14	50.0	36	2	US-08-764-100-6 Sequence 6, Appl
9	13.8	49.3	35	3	US-08-583-276-9 Sequence 9, Appl
10	13.8	49.3	35	3	US-08-583-276-10 Sequence 10, Appl
11	13.6	48.6	21	4	US-08-743-637B-210 Sequence 210, Appl
12	13.6	48.6	26	2	US-08-538-875-26 Sequence 26, Appl
13	13.4	47.9	18	5	US-09-289-376-21 Sequence 21, Appl
14	13.4	47.9	44	2	US-08-592-406-5 Sequence 5, Appl
15	13.4	47.9	97	1	US-07-635-561A-3 Sequence 3, Appl
16	13.2	47.1	23	3	US-08-359-705B-11 Sequence 11, Appl
17	13.2	47.1	23	3	US-08-286-846A-11 Sequence 11, Appl
18	13.2	47.1	23	4	US-08-457-880A-11 Sequence 11, Appl
19	13.2	47.1	23	5	US-08-444-622A-11 Sequence 11, Appl
20	13.2	47.1	23	5	US-08-942-562-11 Sequence 11, Appl
21	13.2	47.1	29	1	US-08-105-483-268 Sequence 268, App
22	13.2	47.1	29	2	US-08-224-391-84 Sequence 84, Appl
23	13.2	47.1	29	2	US-08-484-304-84 Sequence 84, Appl
24	13.2	47.1	29	2	US-08-224-657-105 Sequence 105, App
25	13.2	47.1	29	2	US-08-709-209-268 Sequence 268, App
26	13.2	47.1	29	2	US-08-257-073-63 Sequence 63, Appl
27	13.2	47.1	29	2	US-08-458-101-266 Sequence 268, App

28	13.2	47.1	29	3	US-08-184-009-134	Sequence 134, App
29	13.2	47.1	29	3	US-08-566-398-45	Sequence 45, Appl
30	13.2	47.1	29	4	US-08-458-356-134	Sequence 134, App
31	13.2	47.1	29	4	US-08-658-665-88	Sequence 24, Appl
32	13.2	47.1	43	1	US-07-885-689A-24	Sequence 116, Appl
33	13.2	47.1	43	4	US-08-857-946-116	Sequence 116, App
34	13.2	47.1	43	5	US-08-970-740-116	Sequence 38, Appl
35	13	46.4	21	1	US-08-434-411-38	Sequence 38, Appl
36	13	46.4	21	1	US-08-434-402-38	Sequence 38, Appl
37	13	46.4	21	2	US-08-783-288-38	Sequence 38, Appl
38	13	46.4	21	4	US-08-890-640-38	Sequence 38, Appl
39	13	46.4	21	7	5194592-69	Patent No. 5194592
40	13	46.4	30	6	PCT-US93-01598-27	Sequence 27, Appl
41	13	46.4	48	1	US-08-116-389-16	Sequence 16, Appl
42	13	46.4	48	2	US-08-708-431-16	Sequence 16, Appl
43	13	46.4	48	3	US-08-880-830-16	Sequence 16, Appl
44	13	46.4	48	6	PCT-US94-13895-16	Sequence 16, Appl
45	13	46.4	56	6	PCT-US91-05177-18	Sequence 18, Appl

# ALIGNMENTS

RESULT 1  
Sequence 32, Application US/08361920  
Patent No. 5457046  
GENERAL INFORMATION:  
APPLICANT: Moeldike, Helle F.  
APPLICANT: Hagen, Frederick  
APPLICANT: Hjort, Carsten M.  
APPLICANT: Sven, Hastrup  
TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose  
NUMBER OF SEQUENCES: 85  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NO. 54570460 No. 5457046disk of No. 5457046th America, Inc.  
STREET: 405 Lexington Avenue, 62nd Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08361, 920  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/940, 860  
FILING DATE: 28-OCT-1992  
APPLICATION NUMBER: DK 1158/90  
FILING DATE: 09-MAY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK91/00124  
FILING DATE: 08-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Landitis, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 3435, 204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-867-0298  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 39 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

HYPOTHEICAL: NO  
ANTI-SENSE: NO  
US-08-361-920-32

Query Match 57.1%; Score 16; DB 1; Length 39;  
Best Local Similarity 79.2%; Pred. No. 47;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 4 gaccatttggcaccacatag 27  
||||| ||| | ||| ||  
Db 11 GACCAATTCTGGCCACACATTGG 34

RESULT 2  
US-08-479-939-32

Sequence 32, Application US/08479939  
Patent No. 5686593

GENERAL INFORMATION:

APPLICANT: Woeldike, Helle F.

APPLICANT: Hagen, Frederick

APPLICANT: Hjort, Carsten M.

APPLICANT: Sven, Hastrup

TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose

TITLE OF INVENTION: or Hemicellulose

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 56865930 No. 5686593disk of No. 5686593th America, Inc.

STREET: 405 Lexington Avenue, 62nd floor

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10174-6201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/479, 939

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/361, 920

FILING DATE: 22-DEC-1994

APPLICATION NUMBER: US 07/940, 860

FILING DATE: 28-OCT-1992

APPLICATION NUMBER: DK 1158/90

FILING DATE: 09-MAY-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/DK91/00124

FILING DATE: 08-MAY-1991

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J.

REGISTRATION NUMBER: 33,728

REFERENCE/DOCKET NUMBER: 3435, 204-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123

TELEFAX: 212-867-0298

INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:

LENGTH: 39 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHEICAL: NO

ANTI-SENSE: NO

US-08-479-939-32

Query Match 57.1%; Score 16; DB 1; Length 39;

Best Local Similarity 79.2%; Pred. No. 47;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Oy 4 gaccatttggcaccacatag 27  
||||| ||| | ||| ||  
Db 11 GACCAATTCTGGCCACACATTGG 34

RESULT 3  
US-08-483-432-32

Sequence 32, Application US/08483432  
Patent No. 5763254

GENERAL INFORMATION:

APPLICANT: Woeldike, Helle F.

APPLICANT: Hagen, Frederick

APPLICANT: Hjort, Carsten M.

APPLICANT: Sven, Hastrup

TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose

TITLE OF INVENTION: or Hemicellulose

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 57632540 No. 5763254disk of No. 5763254th America, Inc.

STREET: 405 Lexington Avenue, 62nd floor

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10174-6201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/483, 432

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/361, 920

FILING DATE: 22-DEC-1994

APPLICATION NUMBER: US 07/940, 860

FILING DATE: 28-OCT-1992

APPLICATION NUMBER: DK 1158/90

FILING DATE: 09-MAY-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/DK91/00124

FILING DATE: 08-MAY-1991

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J.

REGISTRATION NUMBER: 33,728

REFERENCE/DOCKET NUMBER: 3435, 204-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123

TELEFAX: 212-867-0298

INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:

LENGTH: 39 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHEICAL: NO

ANTI-SENSE: NO

US-08-483-432-32

Query Match 57.1%; Score 16; DB 2; Length 39;  
Best Local Similarity 79.2%; Pred. No. 47;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 4 gaccatttggcaccacatag 27  
||||| ||| | ||| ||  
Db 11 GACCAATTCTGGCCACACATTGG 34



```

RESULT 4
US-08-459-135A-2/C
; Sequence 2, Application US/08459135A
; Patent No. 5955078
; GENERAL INFORMATION:
; APPLICANT: BURNHAM, Martin Karl Russel
; APPLICANT: CHOPRA, Ian
; APPLICANT: CRITCHLEY, Ian Alfred
; APPLICANT: KNOWLES, David Justin Charles
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN: MONOCLONAL ANTIBODY AND THEIR US
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fastseq, Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,135A
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB/94/00215
; FILING DATE: 04-Feb-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimm1, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P30591C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: GENOMIC DNA
; US-08-459-135A-2

Query Match 53.6%; Score 15; DB 4; Length 34;
Best Local Similarity 78.3%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0;

QY 6 ccaatttgcacacaataaggg 28
||||| ||||||| |||
Db 29 CCAATCGTGCACCAACGTAAGG 7

RESULT 5
US-08-198-431-31
; Sequence 31, Application US/08198431
; Patent No. 5605793
; GENERAL INFORMATION:
; APPLICANT: Stemmer, Willem P.C.
; TITLE OF INVENTION: Methods for In Vitro Recombination
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22313

```

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Floppy disk
3  COMPUTER: IBM PC compatible
4  OPERATING SYSTEM: PC-DOS/MS-DOS
5  SOFTWARE: Patent In Release #1.0, Version #1.25
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: US/08/198,431
8  FILING DATE: 17-FEB-1994
9  CLASSIFICATION: 435
10 ATTORNEY/AGENT INFORMATION:
11 NAME: Wool, Leslie
12 REGISTRATION NUMBER: 37,047
13 REFERENCE/DOCKET NUMBER: 000324-003
14 TELECOMMUNICATION INFORMATION:
15 TELEPHONE: 415-854-7400
16 TELEFAX: 415-854-8275
17 INFORMATION FOR SEQ ID NO: 31:
18 SEQUENCE CHARACTERISTICS:
19 LENGTH: 72 base pairs
20 TYPE: nucleic acid
21 STRANDEDNESS: single
22 TOPOLOGY: linear
23 MOLECULE TYPE: DNA (oligonucleotide)
24 US-08-198-431-31
25
26 Query Match 50.7%; Score 14.2; DB 1; Length 72;
27 Best Local Similarity 70.4%; Pred. No. 3.4e+02;
28 Matches 19; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
29
30 Oy 1 cttgaccaatttggcaccacataag 27
31 | | | | | | | | | | | | | | | | |
32 Db 39 CGTTACATATATTGTGACCAACAGAG 65
33
34 RESULT 6
35 US-08-564-955-31
36 Sequence 31, Application US/08564955
37 Patent No. 5811238
38
39 GENERAL INFORMATION:
40 APPLICANT: STEMMER, WILLEM P.C.
41 APPLICANT: CRAMERI, ANDREAS M.
42 TITLE OF INVENTION: METHODS FOR GENERATING POLYNUCLEOTIDES
43 TITLE OF INVENTION: HAVING DESIRED CHARACTERISTICS BY ITERATIVE SELECTION AND
44 TITLE OF INVENTION: RECOMBINATION
45 NUMBER OF SEQUENCES: 67
46 CORRESPONDENCE ADDRESS:
47 ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW
48 STREET: TWO EMBARCADERO CENTER, 8TH FLOOR
49 CITY: SAN FRANCISCO
50 STATE: CALIFORNIA
51 COUNTRY: U.S.A.
52 ZIP: 94111-3834
53
54 COMPUTER READABLE FORM:
55 MEDIUM TYPE: Floppy disk
56 COMPUTER: IBM PC compatible
57 OPERATING SYSTEM: PC-DOS/MS-DOS
58 SOFTWARE: Patent In Release #1.0, Version #1.25
59 CURRENT APPLICATION DATA:
60 APPLICATION NUMBER: US/08/564,955
61 FILING DATE: 30-NOV-1995
62 CLASSIFICATION: 435
63 PRIOR APPLICATION NUMBER:
64 APPLICATION NUMBER: US 08/198,431
65 FILING DATE: 17-FEB-1994
66 PRIOR APPLICATION DATA:
67 APPLICATION NUMBER: US 08/537,874
68 FILING DATE: 30-OCT-1995
69 PRIOR APPLICATION DATA:
70 APPLICATION NUMBER: PCT/US95/02126
71 FILING DATE: 17-FEB-1995
72 ATTORNEY/AGENT INFORMATION:
73 NAME: DUNN, TRACY J.

```





STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-743-637B-210

Query Match 48.6%; Score 13.6; DB 4; Length 21;  
Best Local Similarity 80.0%; Pred. No. 5.1e+02;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 3 tgaccatttggcaccac 22  
||||| ||||| || |||||  
Db 1 TGACCATTTTATCAGCAAC 20

RESULT 12  
US-08-538-875-26/c  
Sequence 26, Application US/08538875  
Patent No. 5773582  
GENERAL INFORMATION:  
APPLICANT: Shin, Hang-Cheol  
APPLICANT: Shin, Nam-Kyu  
APPLICANT: Lee, Inkyung  
APPLICANT: Kang, Sungzong  
TITLE OF INVENTION: TUMOR NECROSIS FACTOR MUTAINS  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shin, Hang-Cheol  
STREET: Jukong Gochung Apt. 1014-806, Haan-dong  
CITY: Kwangmyung-shi  
STATE: Kyungki-do  
COUNTRY: Republic of Korea  
ZIP: 423-060  
ADDRESSEE: Shin, Nam-Kyu  
STREET: #181-404 Sadang-4-dong, Dongjak-ku  
CITY: Seoul  
STATE:  
COUNTRY: Republic of Korea  
ZIP: 156-094  
ADDRESSEE: Lee, Inkyung  
STREET: 11/2, #302-39 Juan-4-dong, Nam-ku  
CITY: Incheon  
STATE:  
COUNTRY: Republic of Korea  
ZIP: 402-204  
ADDRESSEE: Kang, Sungzong  
STREET: #84-4 Daeshin-dong, Seodaemun-ku  
CITY: Seoul  
STATE:  
COUNTRY: Republic of Korea  
ZIP: 120-160  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette 3.5inch 2.0mb storage  
COMPUTER: IBM PC/AT  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/538, 875  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/193, 336  
FILING DATE:  
APPLICATION NUMBER: KR 93-1751  
FILING DATE: 9-FEB-1993  
ATTORNEY/AGENT INFORMATION:  
NAME:  
REGISTRATION NUMBER:  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE:  
TELEFAX:  
INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:  
LENGTH: 26 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: primer DNA  
US-08-538-875-26

Query Match 48.6%; Score 13.6; DB 2; Length 26;  
Best Local Similarity 80.0%; Pred. No. 5.3e+02;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 4 gaccatttggcaccacaa 23  
||| ||||| ||||| ||  
Db 20 GCCCTTTTGGCACCACCA 1

RESULT 13  
US-09-289-376-21  
Sequence 21, Application US/09289376  
Patent No. 6013788  
GENERAL INFORMATION:  
APPLICANT: Brett P. Monia  
APPLICANT: Lex M. Cowser  
TITLE OF INVENTION: ANTISENSE MODULATION OF SMAD3 EXPRESSION  
FILE REFERENCE: RTS-0043  
CURRENT APPLICATION NUMBER: US/09/289,376  
CURRENT FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 47  
SEQ ID NO 21  
LENGTH: 18  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Antisense Oligonucleotide  
US-09-289-376-21

Query Match 47.9%; Score 13.4; DB 5; Length 18;  
Best Local Similarity 93.3%; Pred. No. 6e+02;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 13 tggcaccacacatag 27  
||||| ||||| ||||  
Db 4 tggcaccacacacag 18

RESULT 14  
US-08-592-406-5/c  
Sequence 5, Application US/08592406  
Patent No. 5821059  
GENERAL INFORMATION:  
APPLICANT: MINION, F. Chris  
APPLICANT: KNUDSON, Kevin L.  
TITLE OF INVENTION: MYCOPLASMA EXPRESSION SYSTEM  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/592,406  
FILING DATE: 06-FEB-1996  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US93/07407  
FILING DATE: 06-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 76645/132  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-592-406-5

Query Match 47.9%; Score 13.4; DB 2; Length 44;  
Best Local Similarity 93.3%; Pred. No. 7.2e+02;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 cttgaccaatttgg 15  
|||||  
DB 33 CTTGACCAATTGGG 19

RESULT 15  
US-07-635-561A-3/C  
Sequence 3, Application US/07635561A  
Patent No. 5244805  
GENERAL INFORMATION:  
APPLICANT: Miller, Lois K.  
TITLE OF INVENTION: Improved Baculovirus Expression Vectors  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee and Winner  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: CO  
COUNTRY: USA  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/635,561A  
FILING DATE: 19910117  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/353,847  
FILING DATE: 17-MAY-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO 90/02814  
FILING DATE: 17-MAY-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 4-90A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/499-8080  
TELEFAX: 303/499-8089  
TELEX: 823189  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 97 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
US-07-635-561A-3

Query Match 47.9%; Score 13.4; DB 1; Length 97;  
Best Local Similarity 73.9%; Pred. No. 8.4e+02;  
Matches 17; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 3 tgaccaatttggcaccacata 25  
|||||  
DB 25 TCACCAAGCTGTGCTCCATGATA 3

Search completed: June 4, 2000, 16:09:42  
Job time: 28061 sec

4 . . 2

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 13:53:12 ; Search time 4521.53 Seconds  
(without alignments)  
25.100 Million cell updates/sec

Title: US-09-164-714-11

Perfect score: 28

Sequence: 1 cttgaccaatttgcacacacataagg 28

Scoring table:

IDENTITY\_NDC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 119968

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result

Query

SUMMARIES

No.	Score	Match	Length	DB	ID	Description
C 1	15.6	55.7	67	34	AA474067	AA474067 ve54e02.x
C 2	15.2	54.3	46	43	AI209047	AI209047 qg28e02.x
C 3	15	53.6	73	51	AI739189	AI739189 w137c11.x
C 4	15	53.6	82	51	AI745436	AI745436 t049t11.x
C 5	15	53.6	89	23	H44598	H44598 yp17c09.r1
C 6	15	53.6	93	44	AI282700	AI282700 q183h07.x
C 7	15	53.6	97	37	AA714476	AA714476 nw06f12.s
C 8	14.8	52.9	95	35	AA566933	AA566933 987 Ldb10
C 9	14.8	52.9	87	60	AI795116	AI795116 sb76b06.y
C 10	14.6	52.1	55	46	AI444123	AI444123 fb26b06.x
C 11	14.6	52.1	59	62	AI880479	AI880479 at78d09.x
C 12	14.6	52.1	92	23	H40392	H40392 yp60e01.r1
C 13	14.6	52.1	99	70	AA145122	AA145122 q429e08.y
C 14	14.4	51.4	79	29	AA165762	AA165762 ms60f11.r
C 15	14.4	51.4	91	81	AA394554	AA394554 sh32e03.y
C 16	14.4	51.4	96	44	AI314653	AI314653 u378t11.x
C 17	14.4	51.4	98	41	AU008426	AU008426 AU008426
C 18	14.2	50.7	58	28	AA107577	AA107577 mp05a11.r
C 19	14.2	50.7	97	29	AA180917	AA180917 zp44b01.r
C 20	14	50.0	46	40	AA95439	AA95439 os84f01.s
C 21	14	50.0	88	47	AI474685	AI474685 tm35f12.x
C 22	14	50.0	91	36	AA659101	AA659101 nu81e01.s
C 23	13.8	49.3	86	29	AA146242	AA146242 mg8f06.r
C 24	13.8	49.3	89	26	W76217	W76217 zd58f09.r1
C 25	13.8	49.3	92	46	AI423670	AI423670 tf85c04.x
C 26	13.8	49.3	99	41	AU007058	AU007058 AU007058
C 27	13.6	48.6	56	84	B03327	B03327 cSR1-176p2-
C 28	13.6	48.6	61	38	AA746401	AA746401 nw62c10.s
C 29	13.6	48.6	63	45	AI338554	AI338554 qg93c08.x
C 30	13.6	48.6	88	24	N25181	N25181 yy02b10.s1
C 31	13.6	48.6	91	34	AA472378	AA472378 v06g906.r
C 32	13.6	48.6	95	40	AA910572	AA910572 ok66h10.s
C 33	13.6	48.6	97	44	AI285065	AI285065 qk56f09.x
C 34	13.6	48.6	98	29	AA187895	AA187895 zp74g12.r
C 35	13.6	48.6	98	50	F29909	F29909 HSPD20053 H
C 36	13.4	47.9	58	41	AU007255	AU007255 AU007255
C 37	13.4	47.9	64	46	AI442743	AI442743 sa85h01.y
C 38	13.4	47.9	67	30	AA232723	AA232723 z175e10.r
C 39	13.4	47.9	67	37	AA669910	AA669910 ag42g07.s
C 40	13.4	47.9	68	37	AA667085	AA667085 v187c12.s
C 41	13.4	47.9	79	33	AA446093	AA446093 zw58a10.s
C 42	13.4	47.9	98	46	AI427907	AI427907 m138e10.s
C 43	13.4	47.9	100	29	AA152082	AA152082 z148q12.r
C 44	13.2	47.1	34	45	AI367088	AI367088 qg46d03.x
C 45	13.2	47.1	49	24	H97554	H97554 yw04h01.s1

## ALIGNMENTS

RESULT 1  
AA474067/c 67 bp mRNA EST 18-JUN-1997  
ve54e02.x1 Beddington mouse embryonic region Mus musculus cDNA  
clone IMAGE:821978 5' similar to TR:c476095 G476095 B4-2 PROTEIN.  
; mRNA sequence.

ACCESSION AA474067  
VERSION AA474067.1 GI:2202294  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 67)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)

## COMMENT

On May 9, 1995 this sequence version replaced g1:803068.  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:490258  
Trace considered overall poor quality  
Possible reversed clone; similarity on wrong strand  
Seq primer: -26m13 rev1 Et from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers

## FEATURES

SOURCE  
1..67  
/organism="Mus musculus"  
/strain="C57Bl6 x DBA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:821978"  
/clone\_lib="Beddington mouse embryonic region"  
/sex="pooled"  
/tissue\_type="embryo"  
/dev\_stage="7.5dpc"  
/lab\_host="DH12S"  
/note="Organ: whole embryo; Vector: PCMV-SPORT; Site:1:  
SalI; Site:2: NotI; Cloned unidirectionally. Primer:  
Oligo dt. Gastrulating embryos were collected at 7.5dpc  
from C57Bl6 x DBA matings, excluding embryos that had  
developed head folds and all extraembryonic tissues.  
Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).  
Referenced in Development 121, 2479-2489 (1995)"

## BASE COUNT

ORIGIN  
19 a 13 g 12 t  
Query Match 55.7%; Score 15.6; DB 34; Length 67;  
Best Local Similarity 81.8%; Pred. No. 3; le+03;  
Matches 18; Conservative 0; Mismatches 4; Indels 0;

QY 5 acccaatttgcacacacatag 26  
11111111111111111111  
Db 25 ACTAATTTGGCTCCACGATAG 4

RESULT 2  
AI209047/c 46 bp mRNA EST 29-NOV-1998  
qg28e02.x1 NCI-CGAP\_K1d3 Homo sapiens cDNA clone IMAGE:1762394 3'  
similar to SW:Y124\_HUMAN Q14137 HYPOTHETICAL PROTEIN KIAA0124 ; ,  
mRNA sequence.

ACCESSION AI209047  
VERSION AI209047.1 GI:3770989  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 46)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Jan 19, 1998 this sequence version replaced gi:2151684.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center



Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: [www.bio.llnl.gov/db/rp/image/image.html](http://www.bio.llnl.gov/db/rp/image/image.html)

Trace considered overall poor quality  
Insert Length: 692 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 1.  
Location/Qualifiers

## FEATURES

source

1. 46

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1762394"

/clone\_lib="NCI\_CGAP\_Kid3"

/lab\_host="DH10B"

/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 10 a 9 c 17 g 10 t  
ORIGIN

## Query Match

Best Local Similarity 54.3%; Score 15.2; DB 43; Length 46;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 cttgaccaatttggcaccac 20  
||| ||| ||| ||| ||| |||

Db 37 CTTTACCACTTTTGCACCA 18

## RESULT 3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 73

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2217380"

/clone\_lib="NCI\_CGAP\_Ov23"

/tissue\_type="tumor, 5 pooled (see description)"

/lab\_host="DH10B"

/note="Organ: ovary; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.35 kb. Tumor types include: mixed Mullerian tumor, papillary serous, clear cell, spindle cell. All are primary tumors, metastasis positive. Life technologies catalog #: 11534-013"

Seq primer: -40UP from Gibco.

Location/Qualifiers

1. 82

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2217380"

/clone\_lib="NCI\_CGAP\_Ov23"

/tissue\_type="tumor, 5 pooled (see description)"

/lab\_host="DH10B"

/note="Organ: ovary; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.35 kb. Tumor types include: mixed Mullerian tumor, papillary serous, clear cell, spindle cell. All are primary tumors, metastasis positive. Life technologies catalog #: 11534-013"

BASE COUNT 17 a 15 c 10 g 40 t

/clone="IMAGE:2391476"  
/clone\_lib="NCI\_CGAP\_Col6"  
/tissue\_type="colon tumor, RER+  
/lab\_host="DH10B"

/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; Plasmid DNA from the normalized library NCI\_CGAP\_Col6 was prepared, and ss circles were made in vitro. Following BAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1057416-1061255, and 1144584-1145351). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 17 a 18 c 13 g 25 t  
ORIGIN

## Query Match

Best Local Similarity 53.6%; Score 15; DB 51; Length 73;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 6 ccaatttggcaccacatagg 28  
||| ||| ||| ||| ||| |||

Db 58 CCAGCTGTGCTACACATAGG 36

## RESULT 4

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 82

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2217380"

/clone\_lib="NCI\_CGAP\_Ov23"

/tissue\_type="tumor, 5 pooled (see description)"

/lab\_host="DH10B"

/note="Organ: ovary; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.35 kb. Tumor types include: mixed Mullerian tumor, papillary serous, clear cell, spindle cell. All are primary tumors, metastasis positive. Life technologies catalog #: 11534-013"

Seq primer: -40UP from Gibco.

Location/Qualifiers

1. 82

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2217380"

/clone\_lib="NCI\_CGAP\_Ov23"

/tissue\_type="tumor, 5 pooled (see description)"

/lab\_host="DH10B"

/note="Organ: ovary; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.35 kb. Tumor types include: mixed Mullerian tumor, papillary serous, clear cell, spindle cell. All are primary tumors, metastasis positive. Life technologies catalog #: 11534-013"

BASE COUNT 17 a 15 c 10 g 40 t

## ORIGIN

Query Match 53.6%; Score 15; DB 51; Length 82;  
Best Local Similarity 78.3%; Pred. No. 6e+03;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 6 ccaatttggcaccacacatagg 28  
||| | ||| ||||| |||  
Db 76 CCAGCTGTGCTACCAACATAGG 54

RESULT 5  
H44598 89 bp mRNA EST 31-JUL-1995  
LOCUS YP17C09.r1 Soares breast 3NBHst Homo sapiens cDNA clone  
DEFINITION IMAGE:187696 5' similar to gb:M80563 PLACENTAL CALCIUM-BINDING  
PROTEIN (HUMAN);, mRNA sequence.

ACCESSION H44598  
VERSION H44598.1 GI:920650  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 89)  
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rfkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevaskis,E., Waterston,R., Williamson,A., Woldmann,P. and  
Wilson,R.

TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

Insert Size: 585  
High quality sequence starts: 1  
High quality sequence stops: 1  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Trace considered overall poor quality  
Insert Length: 585 Std Error: 0.00  
Seq primer: M13rev

High quality sequence stop: 1.  
Location/Qualifiers  
1. 89  
/organism="Homo sapiens"  
/db\_xref="GDB:3818592"  
/db\_xref="taxon:9606"  
/clone="IMAGE:187696"  
/clone\_lib="Soares breast 3NBHst"  
/sex="Female"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: breast; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTACCAATCTGAGTGGAGCGCGCCCTTTTCTTTTCTTTTCTTTT 3']  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of a modified pT73 vector (Pharmacia).  
Library went through one round of normalization to a Cot =  
20. Library constructed by Bento Soares and M.Fatima  
Bonaldo."

BASE COUNT 22 a 19 c 15 g 23 t 10 others  
ORIGIN

Query Match 53.6%; Score 15; DB 23; Length 89;  
Best Local Similarity 72.0%; Pred. No. 6.1e+03;  
Matches 18; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 3 tgaccaatttggcaccacatagg 27  
| | | | | | | | | | | | | | | | |  
Db 14 TTAGCAACTTGTGCACACACAGG 38

RESULT 6  
A1282700/c 93 bp mRNA EST 23-NOV-1998  
LOCUS gtB3h07.x1 NCI-CGAP.Co14 Homo sapiens cDNA clone IMAGE:1961917 3'  
DEFINITION similar to gb:X17206 40S RIBOSOMAL PROTEIN S4 (HUMAN);, mRNA  
sequence.

ACCESSION A1282700  
VERSION A1282700.1 GI:3920933  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 93)  
AUTHORS Eutheria; Primates; Catarrhini; Homiidae; Homo.  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL Unpublished (1997)  
COMMENT On Jan 17, 1998 this sequence version replaced gi:1900073.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: RobertC.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www.bio.lnl.gov/dbp/image/image.html

Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1. 93  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1961917"  
/clone\_lib="NCI-CGAP.Co14"  
/tissue\_type="moderately-differentiated adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.7 kb. Life Technologies catalog #:  
11531-019"

BASE COUNT 20 a 22 c 21 g 30 t  
ORIGIN

Query Match 53.6%; Score 15; DB 44; Length 93;  
Best Local Similarity 78.3%; Pred. No. 6.1e+03;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 6 ccaatttggcaccacacatagg 28  
||| | ||| ||||| |||  
Db 59 CCAGCTGTGCTACCAACATAGG 37

RESULT 7  
AA714476 97 bp mRNA EST 22-JAN-1998  
LOCUS nw06f12.s1 NCI CGAP.SS1 Homo sapiens cDNA clone IMAGE:1238543 3'  
DEFINITION similar to gb:X17206 40S RIBOSOMAL PROTEIN S4 (HUMAN);, mRNA  
sequence.

ACCESSION AA714476







/lab\_host="DH10B"  
/note="vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Construction of the cDNA library was carried out using Stratagene's 'Unizap - cDNA synthesis kit'. cDNA was constructed using an oligo dt primer/linker that contains a XhoI site within it. Following ds cDNA synthesis, EcoRI adapters were ligated to the blunt ends and sample was digested with XhoI. The result is cDNA with an EcoRI sticky end on one side and a XhoI sticky end on the other. This cDNA was ligated directionally in Unizap arms. The vector is designed containing the pBluescript sequence as well as lambda DNA and cDNA is cloned within this pBluescript sequence. The vector was then packaged using Gold gigaPackaging extracts. Library was grown in XL1blue MRF' cells and amplified. The library was excised by mass excision using Stratagene's 'Mass excision kit' that uses exsist as a helper phage that releases the pBluescript sequence and circularises it as single stranded plasmids that are then packaged (by helper phage) and secreted out of the host cell as phagemids. SOLR cells were transformed with phagemids and the library was plated out on LB-amp plates to select for transformants. Approximately 1,000,000 colonies were grown and recovered. The double stranded plasmid library was recovered by using Qiaagen MidI prep kit. 2 micro grams of each library were used to transform DH10B cells by electroporation."

BASE COUNT 27 a 23 c 24 g 25 t  
ORIGIN

Query Match 52.1%; Score 14.6; DB 70; Length 99;  
Best Local Similarity 81.0%; Pred. No. 9.4e+03;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ctggaccaatttgcaccaa 21  
||||| ||||| ||||| |||||  
Db 69 cttggtcattctgtcctcaa 89

RESULT 14  
LOCUS AA165762 79 bp mRNA EST 12-FEB-1997  
DEFINITION ms06f11.r1 StrataGene mouse embryonic carcinoma (#937317) Mus musculus cDNA clone IMAGE:615981 5' similar to TR:E93245 E93245 ETN  
INSERT IN THE FAS APOPTOSIS GENE OF MRL-IPR/IPR. [1] ; mRNA  
SEQUENCE.  
AA165762  
VERSION AA165762.1 GI:1743977  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 79)  
Mastra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.  
TITLE The WashU-HM Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1325676.  
Contact: Marra M/Mouse EST Project  
WashU-HM Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
MG:376805  
Seq primer: -28m13 rev1 ET from Amersham

FEATURES  
SOURCE  
High quality sequence stop: 75.  
Location/Qualifiers  
1..79  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:615981"  
/clone\_11b="Stratagene mouse embryonic carcinoma (#937317)"  
/tissue\_type="carcinoma"  
/dev\_stage="embryonic"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. p19 cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCAGG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3' "

BASE COUNT 20 a 19 c 26 g 14 t  
ORIGIN

Query Match 51.4%; Score 14.4; DB 29; Length 79;  
Best Local Similarity 75.0%; Pred. No. 1.1e+04;  
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 5 accaatltygcacacacataggg 28  
| ||||| ||||| |||||  
Db 33 AACAAATTTTGGCGCCAGAACTGGG 56

RESULT 15  
LOCUS AM394554/C 91 bp mRNA EST 07-FEB-2000  
DEFINITION sh32e03.y1 Gm-c1017 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1017-3413 5', mRNA sequence.  
ACCESSION AM394554  
VERSION AM394554.1 GI:6912966  
KEYWORDS EST.  
SOURCE soybean.  
ORGANISM Glycine max  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.  
1 (bases 1 to 91)  
Shoemaker,R., Kelm,P., Vodka,L., Erpelting,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Persson,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.  
Public Soybean EST Project  
Unpublished (1999)  
On Jul 9, 1999 this sequence version replaced gi:5433560.  
Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com  
Seq primer: -40RP from Gibco.

FEATURES  
SOURCE

Location/Qualifiers  
1..91  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-c1017-3413"  
/clone\_11b="Gm-c1017"  
/tissue\_type="vegetable buds of field grown plants"



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:24:08 ; Search time 322.35 Seconds  
(without alignments)  
21.732 Million cell updates/sec

Title: US-09-164-714-11

Perfect score: 28  
Sequence: 1 cttgacacatttgcacacacataggg 28

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 431286

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15.8	56.4	24	1	X22061
2	15.2	54.3	29	1	V43049
3	15	53.6	34	1	O71835
4	15	53.6	34	1	T18304
5	14.4	51.4	54	1	X27477
6	14.4	51.4	60	1	X27467
7	14.4	51.4	84	1	X27464
8	14.2	50.7	72	1	O97403
9	14.2	50.7	72	1	T73061
10	14.2	50.7	72	1	V54920
11	14	50.0	27	1	V06529
12	13.8	49.3	35	1	O52717
13	13.8	49.3	35	1	O52718
14	13.8	49.3	67	1	T24754
15	13.8	49.3	71	1	V76550
16	13.6	48.6	21	1	V37051
17	13.6	48.6	41	1	X19306
18	13.6	48.6	54	1	V76140
19	13.6	48.6	69	1	V76290
20	13.4	47.9	97	1	O06886
21	13.4	47.9	97	1	O48454
22	13.2	47.1	23	1	T00694
23	13.2	47.1	26	1	V53561
24	13.2	47.1	29	1	O25777
25	13.2	47.1	29	1	O25777
26	13.2	47.1	29	1	O25777
27	13.2	47.1	29	1	O35714
28	13.2	47.1	29	1	O67804
29	13.2	47.1	29	1	O81594
30	13.2	47.1	29	1	T00910
31	13.2	47.1	29	1	T69988
32	13.2	47.1	29	1	T47576
33	13.2	47.1	31	1	X38735
34	13.2	47.1	35	1	V02251
			56	1	V66089

C	35	13.2	47.1	59	1	V77728	Staphylococcus aur
	36	13.2	47.1	72	1	T18145	Oligo used in synt
C	37	13.2	47.1	80	1	O46817	Ovine GHG primer O
C	38	13.2	47.1	94	1	T23702	Human gene signal
	39	13	46.4	21	1	N82281	Fragment encoding
	40	13	46.4	24	1	O78344	Mumps haemagglut
	41	13	46.4	30	1	O03277	Phe-355 mutation o
	42	13	46.4	34	1	T70361	Rev/tx mRNA of bo
C	43	13	46.4	48	1	O87345	Oligonucleotide pr
C	44	13	46.4	48	1	T31534	Test analyte from
C	45	13	46.4	55	1	X15963	PCR primer Al-7 us

## ALIGNMENTS

RESULT	1
X22061	X22061 standard; DNA; 24 BP.
ID	X22061:
AC	19-MAY-1999 (first entry)
DT	PCR primer for C. maltosa cytochrome POX4 gene.
DE	PCR primer; POX4 gene; Pichia pastoris; di-carboxylic acid production;
KW	6-22C mono-carboxylic acid production; alkane hydroxylation; ss.
OS	Synthetic.
PN	Candida maltosa.
PD	WO9904014-A2.
PF	28-JAN-1999.
PR	20-JUL-1998; 014935.
PT	21-JUL-1997; US-053215.
PA	(DUPO) DU POINT DE NEMOURS & CO E. I.
PI	Fallon RD, Payne MS, Picataggio SK, Wu S;
DR	WPI; 99-132258/11.
PT	Production of mono- and di-carboxylic alkanolic acids in new
PT	engineered yeast - containing inserted copies of genes for
PT	cytochrome P450 monooxygenase and/or reductase, providing increased
PT	production and conversion efficiency, useful e.g. as intermediates
PT	for polymers and surfactants
PS	Example 6; Page 33; 57pp; English.
CC	This sequence represents a PCR primer for the POX4 gene
CC	of Candida maltosa. The invention relates to a method for the
CC	production of 6-22C mono- and di-carboxylic acids (I), which comprises
CC	treating, under aerobic conditions: (i) Pichia pastoris having
CC	genetically engineered alkane hydroxylation activity; or (ii) Candida
CC	maltosa genetically engineered to have enhanced alkane hydroxylation
CC	activity and/or a blocked beta-oxidation pathway, with a 6-22C linear
CC	hydrocarbon. The method is specifically used to make dodecanedioic acid
CC	(Ia) from dodecane, but generally any 6-22C alkane, or derived
CC	monocarboxylic acids or esters, can be converted. (I) that are
CC	monocarboxylic acids are intermediates for surfactants and those that are
CC	dicarboxylic acids are intermediates for polymers and anticorrosion
CC	agents. Transformation with the specified genes introduces (i)-producing
CC	ability to P. pastoris and enhances that in C. maltosa, resulting in
CC	conversion efficiency and productivity suitable for a commercial process
CC	(that is less environmentally damaging than known chemical syntheses).
CC	Disruption of the beta-oxidation pathway increases metabolite flow to the
CC	omega-oxidation pathway, and thus yield of, and selectivity for, (I).
SQ	Sequence 24 BP; 6 A; 4 C; 4 G; 10 T;
Query Match	56.4%; Score 15.8; DB 1; Length 24;
Best Local Similarity	89.5%; Pred. No. 1.2e+02;
Matches 17; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
Oy	2 ttgacacatttgcacaca 20
Db	5 ttgacacatttgcacaca 23
RESULT	2
V43049	V43049 standard; DNA; 29 BP.
ID	V43049:
AC	V43049;

DR 23-OCT-1998 (first entry)  
DE 5, PCR primer Deg-A used to amplify nitric oxide synthase DNA.  
KW Nitric oxide synthase; treatment; epidermal; dermal condition;  
KM unwanted pigmentation; PCR primer; ss.  
OS Synthetic.  
OS Mus sp.  
OS Rattus sp.  
OS Homo sapiens.  
PN WO983379-A1.  
PD 06-AUG-1998.  
PF 03-FEB-1998; U01891.  
PR 04-FEB-1997; US-037098.  
PA (GENO ) GEN HOSPITAL CORP.  
PI Lerner EA, Lerner LH, Qureshi AA;  
DR WPI: 98-437056/37.  
PT Treatment of unwanted epidermal or dermal conditions - comprising  
PR administration of treatment which modulates level of nitric oxide in  
PT skin  
PS Disclosure: Page 17, 35pp; English.  
CC PCR primers V43049-50 were used to amplify DNA encoding nitric oxide  
CC synthase (inducible isoform, the isoform found in nerve cells, and  
CC isoform found in epithelial cells). The specification describes  
CC a method for treating a subject for an unwanted epidermal or dermal  
CC condition, which comprises administration of a compound which modulates  
CC the level of nitric oxide in the skin. Where the condition is  
CC characterised by unwanted cells or unwanted pigmentation, the method  
CC includes increasing the level of nitric oxide in the skin. Where the  
CC condition is characterised by the lack of or by an insufficient number  
CC of dermal or epidermal cells or a lack of pigmentation, the method  
CC includes reducing the level of nitric oxide in the skin.  
SQ Sequence 29 Bp; 12 A; 7 C; 5 G; 2 T;

Query Match 54.3%; Score 15.2; DB 1; Length 29;  
Best Local Similarity 77.3%; Pred. No. 2.4e+02;  
Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 7 ccaattggcaccacaatagg 28  
||| | : ||||| ||| |||  
DB 6 CAAGTAYGCCACCAACAAGG 27

RESULT 3  
Q71835/c  
ID Q71835 standard; DNA: 34 Bp.  
AC Q71835;  
DE 30-MAR-1995 (first entry)  
DE Primer F1B2.  
KM Fibronectin binding protein; Fbp; fibronectin binding domain;  
KM monoclonal antibody; Mab; adhesion; Gram-positive bacterium;  
KM bacteremia; Escherichia coli; polymerase chain reaction; PCR;  
KM amplification; primer; Staphylococcus aureus; ss.  
OS Synthetic.  
PN WO9418327-A.  
PD 18-AUG-1994.  
PF 04-FEB-1994; G00215.  
PR 05-FEB-1993; GB-002289.  
PR 20-OCT-1993; GB-021592.  
PA (SMK ) SMITHKLINE BEECHAM PLC.  
PI Burnham MKR, Chopra I, Critchley IA, Knowles DDC;  
DR WPI: 94-279748/34.  
PT Fibronectin binding protein and monoclonal antibodies specific  
PT for Fbp - useful to prevent adherence of Gram-positive bacteria  
PT to indwelling devices or wounds  
PS Disclosure: Page 29, 40pp; English.  
CC Polypeptides corresponding to residues G709-T886 plus P1VPPPT,  
CC G709-P838(P838T) and G709-P838 (R8805-07, respectively) of  
CC S. aureus J2385 (NCIMB 40532) Fbp type A fibronectin binding  
CC domain D1-D4 region (R5808) were expressed in E. coli BL21(DE4)  
CC and used to raise Mabs specific for Fbp. DNA encoding  
CC D1-D4(G709-T886) was obtained by PCR amplification of S. aureus  
CC chromosomal DNA using primers F1B1 (Q71834) and F1B2 (Q71835).  
SQ Sequence 34 Bp; 5 A; 6 C; 14 G; 9 T;

Query Match 53.6%; Score 15; DB 1; Length 34;  
Best Local Similarity 78.3%; Pred. No. 3e+02;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 6 ccaattggcaccacaatagg 28  
|||| | ||||| ||| |||  
DB 29 CCAATGTCGCACCAAGTAAAG 7

RESULT 4  
T18304/c  
ID T18304 standard; DNA: 34 Bp.  
AC T18304;  
DE 13-NOV-1996 (first entry)  
DE Primer for amplifying fibronectin binding domains.  
KM Fibronectin; oral hygiene; gingivitis; gum disease; adhesion;  
KM periodontitis; oral pathogen; caries; calculus; candidiasis;  
KM oral surgery; tissue regeneration; irrigation; toothpaste;  
KM dentifrice; mouthwash; lozenge; ss.  
OS Synthetic.  
PN WO9604003-A1.  
PD 15-FEB-1996.  
PF 18-JUL-1995; E02825.  
PR 05-AUG-1994; GB-015902.  
PA (SMK ) SMITHKLINE BEECHAM PLC.  
PI Barnett P, Critchley IA, Dodd I;  
DR WPI: 96-129122/13.  
PT Prevention of adherence of oral pathogens in the oral cavity,  
PT partic. tooth surfaces - by application of a fibronectin binding  
PT protein or polypeptide or a monoclonal antibody or fragment against  
PT it  
PS Example 1: Page 18, 41pp; English.  
CC Adherence of oral pathogens, particularly to tooth surfaces, can be  
CC prevented by application of a fibronectin binding protein or  
CC polypeptide. The fibronectin binding protein or polypeptide is  
CC useful in the manufacture of oral hygiene compositions, eg.  
CC toothpaste, liquid dentifrice, mouthwash or lozenge. They are  
CC useful to prevent the plaque-related development of carious lesions,  
CC gingivitis, calculus or periodontal disease and to combat oral cavity  
CC infections, e.g. candidiasis. They may also be useful in oral  
CC surgery e.g. in guided tissue regeneration procedures to prevent  
CC subsequent bacterial infection, and for irrigation of periodontal  
CC pockets. Two primers (T18303, T18304) were used to amplify DNA  
CC encoding the fibronectin binding domains of fibronectin binding  
CC protein from the chromosomal DNA of Staphylococcus aureus J2385.  
SQ Sequence 34 Bp; 5 A; 6 C; 14 G; 9 T;

Query Match 53.6%; Score 15; DB 1; Length 34;  
Best Local Similarity 78.3%; Pred. No. 3e+02;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 6 ccaattggcaccacaatagg 28  
|||| | ||||| ||| |||  
DB 29 CCAATGTCGCACCAAGTAAAG 7

RESULT 5  
X27477/c  
ID X27477 standard; DNA: 54 Bp.  
AC X27477;  
DE 08-JUN-1999 (first entry)  
DE Oligo 1131 for hCG alpha-subunit analogues.  
KW Analogue; heterodimeric; glycoprotein hormone; hCG; hLH; hFSH; hTSH;  
KW human chorionic gonadotropin; human luteinising hormone; disulphide bond;  
KW human follicle stimulating hormone; human thyroid stimulating hormone;  
KW stability; primer; amplification; PCR; mutation; ss.  
OS Synthetic.  
OS Homo sapiens.  
PN WO9658957-A2.  
PD 30-DEC-1998.

PF 25-JUN-1998; U13070.  
PR (ISTF ) ARS APPLIED RES SYSTEMS HOLDING NV.  
PA (MCIN-) MCINNIS P G.  
PI Moyle WR;  
DR WPI: 99-081219/07.  
PT New stabilised glycoprotein hormones - particularly hCG, hLH, hFSH  
PT or hTSH, have an intersubunit disulphide crosslink between the  
PS alpha- and beta-subunits to improve stability  
PS Example 1; Page 60; 139pp; English.  
CC The invention relates to the production of analogues of a heterodimeric  
CC subunit glycoprotein hormone (GPH) e.g. human chorionic gonadotropin  
CC (hCG), human luteinising hormone (hLH), human follicle stimulating  
CC hormone (hFSH), human thyroid stimulating hormone (hTSH), and functional  
CC mutants, which are modified to contain an intersubunit disulphide bond,  
CC between an alpha-subunit cysteine and a beta-subunit cysteine, for  
CC improved stability, the analogue retaining at least a portion of the  
CC bioactivity for the corresponding native GPH receptor. Primers  
CC X27450-X27484 were used to introduce the mutations in the hCG sequence.  
CC The improved analogues are designed specifically to reduce perturbation  
CC of the 3-dimensional structure of the hormone, thereby creating greater  
CC likelihood that the dimer will be formed in vivo and the formed dimer  
CC will have affinity for the native receptors and have agonistic activity  
CC on them. The changes stabilise the GPHs and prolong the biological  
CC activities of the hormones. The analogues can have uses as for the  
CC native GPHs.  
SQ Sequence 34 BP; 18 A; 11 C; 12 G; 13 T;  
Query Match 51.4%; Score 14.4; DB 1; Length 54;  
Best Local Similarity 75.0%; Pred. No. 5.9e+02;  
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 3 tgaccatttggcaccacacatag 26  
DB 43 TGACGTCCTTTGGACCAACATAG 20  
RESULT 6  
X27467/C  
ID X27467 standard; DNA; 60 BP.  
AC X27467;  
DR 08-JUN-1999 (first entry)  
DE Oligo 850 for hCG alpha-subunit analogues.  
KW Analogue: heterodimeric; glycoprotein hormone; hCG; hLH; hFSH; hTSH;  
KW human chorionic gonadotropin; human luteinising hormone; disulphide bond;  
KW human follicle stimulating hormone; human thyroid stimulating hormone;  
KW stability; primer; amplification; PCR; mutation; ss.  
OS Synthetic.  
OS Homo sapiens.  
PN WO9858957-A2.  
PD 30-DEC-1998.  
PF 25-JUN-1998; U13070.  
PR 25-JUN-1997; US-050784.  
PA (ISTF ) ARS APPLIED RES SYSTEMS HOLDING NV.  
PA (MCIN-) MCINNIS P G.  
PI Moyle WR;  
DR WPI: 99-081219/07.  
PT New stabilised glycoprotein hormones - particularly hCG, hLH, hFSH  
PT or hTSH, have an intersubunit disulphide crosslink between the  
PS alpha- and beta-subunits to improve stability  
PS Example 1; Page 60; 139pp; English.  
CC The invention relates to the production of analogues of a heterodimeric  
CC subunit glycoprotein hormone (GPH) e.g. human chorionic gonadotropin  
CC (hCG), human luteinising hormone (hLH), human follicle stimulating  
CC hormone (hFSH), human thyroid stimulating hormone (hTSH), and functional  
CC mutants, which are modified to contain an intersubunit disulphide bond,  
CC between an alpha-subunit cysteine and a beta-subunit cysteine, for  
CC improved stability, the analogue retaining at least a portion of the  
CC bioactivity for the corresponding native GPH receptor. Primers  
CC X27450-X27484 were used to introduce the mutations in the hCG sequence.  
CC The improved analogues are designed specifically to reduce perturbation  
CC of the 3-dimensional structure of the hormone, thereby creating greater

CC likelihood that the dimer will be formed in vivo and the formed dimer  
CC will have affinity for the native receptors and have agonistic activity  
CC on them. The changes stabilise the GPHs and prolong the biological  
CC activities of the hormones. The analogues can have uses as for the  
CC native GPHs.  
SQ Sequence 60 BP; 18 A; 13 C; 14 G; 15 T;  
Query Match 51.4%; Score 14.4; DB 1; Length 60;  
Best Local Similarity 75.0%; Pred. No. 6e+02;  
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 3 tgaccatttggcaccacacatag 26  
DB 43 TGACGTCCTTTGGACCAACATAG 20  
RESULT 7  
X27464/C  
ID X27464 standard; DNA; 84 BP.  
AC X27464;  
DR 08-JUN-1999 (first entry)  
DE Oligo 839 for hCG alpha-subunit analogues.  
KW Analogue: heterodimeric; glycoprotein hormone; hCG; hLH; hFSH; hTSH;  
KW human chorionic gonadotropin; human luteinising hormone; disulphide bond;  
KW human follicle stimulating hormone; human thyroid stimulating hormone;  
KW stability; primer; amplification; PCR; mutation; ss.  
OS Synthetic.  
OS Homo sapiens.  
PN WO9858957-A2.  
PD 30-DEC-1998.  
PF 25-JUN-1998; U13070.  
PR 25-JUN-1997; US-050784.  
PA (ISTF ) ARS APPLIED RES SYSTEMS HOLDING NV.  
PA (MCIN-) MCINNIS P G.  
PI Moyle WR;  
DR WPI: 99-081219/07.  
PT New stabilised glycoprotein hormones - particularly hCG, hLH, hFSH  
PT or hTSH, have an intersubunit disulphide crosslink between the  
PS alpha- and beta-subunits to improve stability  
PS Example 1; Page 60; 139pp; English.  
CC The invention relates to the production of analogues of a heterodimeric  
CC subunit glycoprotein hormone (GPH) e.g. human chorionic gonadotropin  
CC (hCG), human luteinising hormone (hLH), human follicle stimulating  
CC hormone (hFSH), human thyroid stimulating hormone (hTSH), and functional  
CC mutants, which are modified to contain an intersubunit disulphide bond,  
CC between an alpha-subunit cysteine and a beta-subunit cysteine, for  
CC improved stability, the analogue retaining at least a portion of the  
CC bioactivity for the corresponding native GPH receptor. Primers  
CC X27450-X27484 were used to introduce the mutations in the hCG sequence.  
CC The improved analogues are designed specifically to reduce perturbation  
CC of the 3-dimensional structure of the hormone, thereby creating greater  
CC likelihood that the dimer will be formed in vivo and the formed dimer  
CC will have affinity for the native receptors and have agonistic activity  
CC on them. The changes stabilise the GPHs and prolong the biological  
CC activities of the hormones. The analogues can have uses as for the  
CC native GPHs.  
SQ Sequence 84 BP; 25 A; 20 C; 16 G; 23 T;  
Query Match 51.4%; Score 14.4; DB 1; Length 84;  
Best Local Similarity 75.0%; Pred. No. 6.3e+02;  
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 3 tgaccatttggcaccacacatag 26  
DB 67 TGACTGCTTTTGGACCAACATAG 44  
RESULT 8  
ID 097403  
AC 097403 standard; DNA; 72 BP.





CC 185 mutation of the mdrl gene and will differ from the sequence of  
 CC GIMDI only at the point mutation wherein codon 185 is changed from  
 CC GGT to GGT. This difference corrects the point mutation in the cDNA  
 CC sequence and encodes Gly in the final construction. The third set of  
 CC primers flank the BstXI and MscI restriction sites. These primers  
 CC 5'Bst and 3'Msc (Q52719, Q52720) overlap the BstXI and MscI sites,  
 CC respectively, and also incorporate another flanking restriction site  
 CC that will allow the PCR fragment to be conveniently subcloned.  
 CC Using the technique of recombinant PCR three separate PCR reactions  
 CC will be initiated, using MD1NA as the template DNA.  
 SQ Sequence 35 BP; 11 A; 8 C; 3 G; 13 T;

Query Match 49.3%; Score 13.8; DB 1; Length 35;  
 Best Local Similarity 88.2%; Pred. No. 1e+03;  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 caatttggcaccaca 23  
 DB 1 CAAATTTGTCACCAATA 17

# RESULT 14

ID T24754 standard; cDNA to mRNA; 67 BP.  
 AC T24754;  
 DT 09-OCT-1996 (first entry)  
 DE Human gene signature HUMGS06829.  
 KM Gene signature: messenger RNA; mRNA; relative abundance; frequency;  
 KW human; cloning; mapping; non-biased library; diagnosis; detection;  
 KW cell typing; abnormal cell function; ss.  
 OS Homo sapiens.  
 PN W09514772-A1.  
 PD 01-JUN-1995.  
 PE 11-NOV-1994; J01916.  
 PR 12-NOV-1993; JP-355504.  
 PA (MATS/) MATSUBARA K.  
 PA (OKUB/) OKUBO K.  
 PI Matsubara K, Okubo K;  
 DR WPI; 95-206931/27.  
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
 PT for diagnosis of abnormal cell function, by preparing cDNA that  
 PT reflects relative abundance of corresp. mRNA in specific human  
 PT tissues  
 PS Claim 1; Page 1687; 2245pp; Japanese.  
 CC A single-stranded DNA (or its complementary strand or the corresp.  
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
 CC given in T19001-T26837 and which is able to hybridise to part of  
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
 CC sequences were obtained from 3'-directed cDNA libraries prepared  
 CC from various human tissues; synthesis of cDNA was initiated from the  
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
 CC untranslated sequence is unique to a particular mRNA species, almost  
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
 CC is constructed so as to reflect accurately the relative abundance of  
 CC different mRNAs in the particular tissue from which it was derived.  
 CC The appearance frequency of a given GS in a cDNA library can be  
 CC determined (esp. using primers and probes derived from the GS  
 CC sequences) as a means of diagnosing abnormal cell function or for  
 CC recognising different cell types.  
 SQ Sequence 67 BP; 26 A; 13 C; 7 G; 20 T;

Query Match 49.3%; Score 13.8; DB 1; Length 67;  
 Best Local Similarity 88.2%; Pred. No. 1e+03;  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 accaatlttgaccaca 21  
 DB 51 ACCAATTTTGCCACAAA 67

# RESULT 15

V76550  
 ID V76550 standard; DNA; 71 BP.  
 AC V76550;  
 DT 16-MAR-1999 (first entry)  
 DE Staphylococcus aureus contig SRQ ID #2239.  
 KM Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.  
 OS Staphylococcus aureus.  
 PN EP-786519-A2.  
 PD 30-JUL-1997.  
 PE 07-JAN-1997; 100117.  
 PF 05-JAN-1996; US-009861.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
 PI Rosen CA;  
 DR WPI; 97-37492/35.  
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -  
 PT stored on computer readable medium and used in the production of  
 PT anti-S.aureus vaccines  
 PS Claim 1; Page 2185; 3271pp; English.  
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S.aureus DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against S.aureus infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S.aureus DNA sequences contained on the  
 CC computer readable medium.  
 SQ Sequence 71 BP; 21 A; 21 C; 14 G; 15 T;

Query Match 49.3%; Score 13.8; DB 1; Length 71;  
 Best Local Similarity 72.0%; Pred. No. 1e+03;  
 Matches 18; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 cttagcaatttggcaccaca 25  
 DB 13 CTTGACCAATAATGTGATGACACATA 37

Search completed: June 4, 2000, 16:24:08  
 Job time: 28904 sec



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:09:42 ; Search time 244.64 Seconds  
(without alignments)  
9.564 Million cell updates/sec

Title: US-09-164-714-12

Perfect score: 18

Sequence: 1 ggcagacaaccacgcctag 18

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 375680

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :  
1: /cgn2\_6/prodata/2/lna/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/2/lna/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/2/lna/5C.COMB.seq:\*  
4: /cgn2\_6/prodata/2/lna/5D.COMB.seq:\*  
5: /cgn2\_6/prodata/2/lna/6.COMB.seq:\*  
6: /cgn2\_6/prodata/2/lna/PCFUS.COMB.seq:\*  
7: /cgn2\_6/prodata/2/lna/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13.8	76.7	25	1	US-08-471-206-10
2	13.8	76.7	25	2	US-08-281-025D-2
3	13.8	76.7	30	1	US-08-471-206-27
4	13.8	76.7	31	1	US-08-471-206-20
5	13.8	76.7	31	1	US-08-471-206-21
6	13.8	76.7	31	2	US-08-281-025D-14
7	13.8	76.7	31	2	US-08-281-025D-15
8	13.8	76.7	50	1	US-08-471-206-26
9	12.2	67.8	50	2	US-08-281-025D-18
10	12.2	67.8	50	2	US-08-281-025D-19
11	12.2	67.8	57	7	5221624-13
12	12.2	67.8	57	7	5221624-15
13	12.2	67.8	57	7	5221624-17
14	12.2	67.8	57	7	5221624-19
15	12.2	67.8	63	7	5221624-23
16	12.2	67.8	64	3	US-08-244-434-5
17	12.2	67.8	66	7	5221624-25
18	12.2	67.8	69	2	US-08-244-492A-8
19	12.2	67.8	69	2	US-08-709-915-10
20	12.2	67.8	75	6	PCT-US94-09653A-32
21	12.2	67.8	84	3	US-08-369-829A-15
22	12.2	67.8	84	4	US-08-386-676E-20
23	12.2	67.8	84	6	PCT-US94-09653A-33
24	11.8	65.6	23	4	US-08-315-695-26
25	11.8	65.6	23	4	US-08-793-410-14
26	11.8	65.6	89	2	US-08-290-592E-38
27	11.8	65.6	89	6	PCT-US96-09448-38

#### ALIGNMENTS

c	28	11.6	64.4	21	3	US-08-468-819-35	Sequence 35, Appl
c	29	11.6	64.4	21	3	US-08-468-819-59	Sequence 59, Appl
c	30	11.6	64.4	42	2	US-08-464-136-46	Sequence 46, Appl
c	31	11.6	64.4	42	2	US-08-440-787A-45	Sequence 45, Appl
c	32	11.6	64.4	42	3	US-08-349-131-46	Sequence 46, Appl
c	33	11.6	64.4	42	5	US-08-470-297A-46	Sequence 45, Appl
c	34	11.6	64.4	42	6	PCT-US91-07141-45	Sequence 45, Appl
c	35	11.6	64.4	42	6	PCT-US91-07149-46	Sequence 46, Appl
c	36	11.4	63.3	20	4	US-09-289-368-81	Sequence 81, Appl
c	37	11.4	63.3	23	1	US-08-433-126A-135	Sequence 135, App
c	38	11.4	63.3	23	2	US-08-433-124A-135	Sequence 135, App
c	39	11.4	63.3	23	6	PCT-US96-06059-135	Sequence 135, App
c	40	11.4	63.3	25	5	US-09-071-101-6	Sequence 6, Appl
c	41	11.4	63.3	30	1	US-08-433-126A-136	Sequence 136, App
c	42	11.4	63.3	30	2	US-08-433-124A-136	Sequence 136, App
c	43	11.4	63.3	30	6	PCT-US96-06059-136	Sequence 136, App
c	44	11.4	63.3	31	1	US-08-433-126A-137	Sequence 137, App
c	45	11.4	63.3	31	2	US-08-433-124A-137	Sequence 137, App

RESULT 1  
US-08-471-206-10/c  
Sequence 10, Application US/08471206  
Patent No. 5683893  
GENERAL INFORMATION:  
APPLICANT: OHI, HIDEYUKI  
APPLICANT: MIURA, MASAMI  
APPLICANT: UNO, SHUSHI  
APPLICANT: CHUGANITSU, MASAKO  
APPLICANT: HIRUMATSU, RYUJI  
APPLICANT: OHMURA, TAKAO  
TITLE OF INVENTION: MUTANT AOX2 PROMOTER, VECTOR CARRYING  
TITLE OF INVENTION: THE SAME, TRANSFORMANT, AND PRODUCTION OF HETEROLOGOUS  
TITLE OF INVENTION: PROTEIN  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS  
STREET: 2100 PENNSYLVANIA AVE., NW  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,206  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/143,158  
FILING DATE: 29-OCT-1993  
APPLICATION NUMBER: JP 215306/1993  
FILING DATE: 06-AUG-1993  
PRIOR APPLICATION DATA: JP 293315/1992  
FILING DATE: 30-OCT-1992  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 293-7060  
TELEFAX: (202) 293-7860  
TELEX: 6491103  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

US-08-471-206-10

Query Match 76.7%; Score 13.8; DB 1; Length 25;  
Best Local Similarity 88.2%; Pred. No. 52;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ggcagcaaacacgccta 17  
|||||  
DB 19 GCGACAAAATAGCCTA 3

RESULT 2

US-08-281-025D-2/c

; Sequence 2, Application US/08281025D

; Patent No. 5707827

; GENERAL INFORMATION:

; APPLICANT: OHI, Hideyuki

; APPLICANT: Mura, Masami

; APPLICANT: Hirumatsu, Ryuji

; APPLICANT: Ohmura, Takao

; TITLE OF INVENTION: MUTANT AOX2 PROMOTER, VECTOR CARRYING

; TITLE OF INVENTION: SAME, TRANSFORMANT, AND PRODUCTION OF HETEROLOGOUS PROTEIN

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK &amp; SEAS

; STREET: 2100 Pennsylvania Avenue, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/281.025D

; FILING DATE: 27-JUL-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 185003/1993

; FILING DATE: 27-JUL-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Nakamura, Dean H.

; REGISTRATION NUMBER: 33,981

; REFERENCE/DOCKET NUMBER: O-35978

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)293-7060

; TELEFAX: (202)293-7860

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 25 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: other nucleic acid

; DESCRIPTION: /desc = "Plasmid DNA"

US-08-281-025D-2

Query Match 76.7%; Score 13.8; DB 2; Length 25;  
Best Local Similarity 88.2%; Pred. No. 52;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ggcagcaaacacgccta 17  
|||||  
DB 19 GCGACAAAATAGCCTA 3

RESULT 3

US-08-471-206-27/c

; Sequence 27, Application US/08471206

; Patent No. 5683893

; GENERAL INFORMATION:

; APPLICANT: OHI, HIDEYUKI

; APPLICANT: MURA, MASAMI

; APPLICANT: UNO, SHUSEI

; APPLICANT: CHUGANUJI, MASAKO

; APPLICANT: HIRUMATSU, RYUJI

; APPLICANT: OHMURA, TAKAO

; TITLE OF INVENTION: MUTANT AOX2 PROMOTER, VECTOR CARRYING

; TITLE OF INVENTION: THE SAME, TRANSFORMANT, AND PRODUCTION OF HETEROLOGOUS

; TITLE OF INVENTION: PROTEIN

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK &amp; SEAS

; STREET: 2100 PENNSYLVANIA AVE., NW

; CITY: WASHINGTON

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/471.206

; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/143,158

; FILING DATE: 29-OCT-1993

; APPLICATION NUMBER: JP 215306/1993

; FILING DATE: 06-AUG-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 293315/1992

; FILING DATE: 30-OCT-1992

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 293-7060

; TELEFAX: (202) 293-7860

; TELEX: 6491103

; INFORMATION FOR SEQ ID NO: 27:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 30 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; US-08-471-206-27

Query Match 76.7%; Score 13.8; DB 1; Length 30;  
Best Local Similarity 88.2%; Pred. No. 53;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ggcagcaaacacgccta 17  
|||||  
DB 23 GCGACAAAATAGCCTA 7

RESULT 4

US-08-471-206-20/c

; Sequence 20, Application US/08471206

; Patent No. 5683893

; GENERAL INFORMATION:

; APPLICANT: OHI, HIDEYUKI

; APPLICANT: MURA, MASAMI

; APPLICANT: UNO, SHUSEI

; APPLICANT: CHUGANUJI, MASAKO

; APPLICANT: HIRUMATSU, RYUJI

; APPLICANT: OHMURA, TAKAO

; TITLE OF INVENTION: MUTANT AOX2 PROMOTER, VECTOR CARRYING

; TITLE OF INVENTION: THE SAME, TRANSFORMANT, AND PRODUCTION OF HETEROLOGOUS

; TITLE OF INVENTION: PROTEIN

NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS  
STREET: 2100 PENNSYLVANIA AVE., NW  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,206  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/143,158  
FILING DATE: 29-OCT-1993  
APPLICATION NUMBER: JP 215306/1993  
FILING DATE: 06-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 293315/1992  
FILING DATE: 30-OCT-1992  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (402) 293-7060  
TELEFAX: (402) 293-7860  
TELEX: 6491103  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Other nucleic acid  
DESCRIPTION: Synthetic DNA  
US-08-471-206-20

```

Query Match      76.7%:  Score 13.8:  DB 1;  Length 31;
Best Local Similarity 88.2%:
Matches 15;  Conservative 0;  Mismatches 2;  Indels 0;  Gaps 0;

OY      1  gcgacaaaccagccta 17
          |||||  |||||
Db      24  GCGACAAAATAGCCTA 8

RESULT      5
US-08-471-206-21
; Sequence 21, Application US/08471206
; Patent No. 5683893
; GENERAL INFORMATION:
; APPLICANT: OHI, HIDEYUKI
; APPLICANT: MIURA, MASAMI
; APPLICANT: UNO, SHUSEI
; APPLICANT: CHUGANUJI, MASAKO
; APPLICANT: HIRUMATSU, RYUJI
; APPLICANT: OHMURA, TAKAO
; TITLE OF INVENTION: MUTANT AOX2 PROMOTER, VECTOR CARRYING
; TITLE OF INVENTION: THE SAME, TRANSFORMANT, AND PRODUCTION OF HETEROLOGOUS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MIOW, ZINN, MACPEAK & SEAS
; STREET: 2100 PENNSYLVANIA AVE., NW
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
;

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,206
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/143,158
FILING DATE: 29-OCT-1993
APPLICATION NUMBER: JP 215306/1993
FILING DATE: 06-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 293315/1992
FILING DATE: 30-OCT-1992
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: Synthetic DNA

```

Query Match	76.7%;	Score 13.8;	DB 1;	Length 31;
Best Local Similarity	88.2%;	Pred. No. 53;		
Matches 15; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

```

QY      1  gcgacaaaaccagccta 17
          |||||
Db      12  GCGACAAAATAGCCTA 28

```

RESULT 6  
US-08-281-025D-14/C  
Sequence 14, Application US/08281025D  
Patent No. 5/07827  
GENERAL INFORMATION:  
APPLICANT: Ohi, Hideyuki  
APPLICANT: Miura, Masami  
APPLICANT: Hiramatsu, Ryuji  
APPLICANT: Ohmura, Takao  
TITLE OF INVENTION: MOTANT AOX2 PROMOTER, VECTOR CARRYING  
TITLE OF INVENTION: SAME, TRANSFORMANT, AND PRODUCTION OF HETEROLOGOUS PROTEIN  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS  
STREET: 2100 Pennsylvania Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/281,025D  
FILING DATE: 27-JUL-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 185003/1993  
FILING DATE: 27-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Nakamura, Dean H.  
REGISTRATION NUMBER: 33,981

REFERENCE/DOCKET NUMBER: Q-35978  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)293-7060  
TELEFAX: (202)293-7860  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Primer"  
US-08-281-025D-14

Query Match 76.7%; Score 13.8; DB 2; Length 31;  
Best Local Similarity 88.2%; Pred. No. 53;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gcgacaaaccagccta 17  
||||||| |||||  
Db 24 GCGACAAATAAGCCTA 8

RESULT 7  
US-08-281-025D-15  
Sequence 15, Application US/08281025D  
Patent No. 5707827  
GENERAL INFORMATION:  
APPLICANT: Ohl, Hideyuki  
APPLICANT: Miura, Masami  
APPLICANT: Hiramatsu, Ryuji  
APPLICANT: Ohmura, Takao  
TITLE OF INVENTION: MUTANT AOX2 PROMOTER, VECTOR CARRYING  
TITLE OF INVENTION: SAME, TRANSFORMANT, AND PRODUCTION OF HETEROLOGOUS PROTEIN  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SUGHRUE, MIOW, ZINN, MACPHEAK & SEAS  
STREET: 2100 Pennsylvania Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/281,025D  
FILING DATE: 27-JUL-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 185003/1993  
FILING DATE: 27-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Nakamura, Dean H.  
REGISTRATION NUMBER: 33,981  
REFERENCE/DOCKET NUMBER: Q-35978  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)293-7060  
TELEFAX: (202)293-7860  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Primer"  
US-08-281-025D-15

Query Match 76.7%; Score 13.8; DB 2; Length 31;  
Best Local Similarity 88.2%; Pred. No. 53;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 gcgacaaaccagccta 17  
||||||| |||||  
Db 12 GCGACAAATAAGCCTA 28

RESULT 8  
US-08-471-206-26/C  
Sequence 26, Application US/08471206  
Patent No. 5683893  
GENERAL INFORMATION:  
APPLICANT: OHI, HIDEYUKI  
APPLICANT: MIURA, MASAMI  
APPLICANT: UNO, SHUSEI  
APPLICANT: CHOGANUJI, MASAKO  
APPLICANT: OHMURA, TAKAO  
TITLE OF INVENTION: MUTANT AOX2 PROMOTER, VECTOR CARRYING  
TITLE OF INVENTION: THE SAME, TRANSFORMANT, AND PRODUCTION OF HETEROLOGOUS  
TITLE OF INVENTION: PROTEIN  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SUGHRUE, MIOW, ZINN, MACPHEAK & SEAS  
STREET: 2100 PENNSYLVANIA AVE., NW  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,206  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/143,158  
FILING DATE: 29-OCT-1993  
APPLICATION NUMBER: JP 215306/1993  
FILING DATE: 06-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 293315/1992  
FILING DATE: 30-OCT-1992  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 293-7060  
TELEFAX: (202) 293-7860  
TELEX: 6491103  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-471-206-26

Query Match 76.7%; Score 13.8; DB 1; Length 50;  
Best Local Similarity 88.2%; Pred. No. 56;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 gcgacaaaccagccta 17  
||||||| |||||  
Db 41 GCGACAAATAAGCCTA 25

RESULT 9  
US-08-281-025D-18/C

```
; Sequence 18, Application US/08281025D
; Patent No. 5707827
;
; GENERAL INFORMATION:
; APPLICANT: Ohl, Hideyuki
; APPLICANT: Miura, Masami
; APPLICANT: Hirata, Ryuji
; APPLICANT: Ohmura, Takao
; TITLE OF INVENTION: MUTANT AOX2 PROMOTER, VECTOR CARRYING
; TITLE OF INVENTION: SAME, TRANSFORMANT, AND PRODUCTION OF HETEROLOGOUS PROTEIN
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/281,025D
; FILING DATE: 27-JUL-1994
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 185003/1993
; FILING DATE: 27-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakamura, Dean H.
; REGISTRATION NUMBER: 33,981
; REFERENCE/DOCKET NUMBER: Q-35978
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Primer"
;
; US-08-281-025D-18

Query Match          67.8%; Score 12.2; DB 2; Length 50;
Best Local Similarity 82.4%; Pred. No. 3.8e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggcacaaacagccta 17
    |||||
DB 41 GGCACAAAATAGCCCA 25

RESULT 10
US-08-281-025D-19/c
; Sequence 19, Application US/08281025D
; Patent No. 5707827
;
; GENERAL INFORMATION:
; APPLICANT: Ohl, Hideyuki
; APPLICANT: Miura, Masami
; APPLICANT: Hirata, Ryuji
; APPLICANT: Ohmura, Takao
; TITLE OF INVENTION: MUTANT AOX2 PROMOTER, VECTOR CARRYING
; TITLE OF INVENTION: SAME, TRANSFORMANT, AND PRODUCTION OF HETEROLOGOUS PROTEIN
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
```

```
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/281,025D
; FILING DATE: 27-JUL-1994
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 185003/1993
; FILING DATE: 27-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakamura, Dean H.
; REGISTRATION NUMBER: 33,981
; REFERENCE/DOCKET NUMBER: Q-35978
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Primer"
;
; US-08-281-025D-19

Query Match          67.8%; Score 12.2; DB 2; Length 50;
Best Local Similarity 82.4%; Pred. No. 3.8e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggcacaaacagccta 17
    |||||
DB 41 GGCACAAAATAGCCCA 25
```

```
RESULT 11
5221624-13/c
; Patent No. 5221624
; APPLICANT: BLAIR, LINDLEY C.; KODURI, JAR-HOW; WEICKMANN,
; JOACHIM J.
; TITLE OF INVENTION: DNA ENCODING (LYS46, ASP97, ASP113) AND
; (LYS46, ASP113, ASP137) THAUMATIN I
; NUMBER OF SEQUENCES: 31
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/407,416
; FILING DATE: 14-SEP-1989
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 268,702
; FILING DATE: 08-NOV-1988
; SEQ ID NO:13
; LENGTH: 57
;
5221624-13

Query Match          67.8%; Score 12.2; DB 7; Length 57;
Best Local Similarity 82.4%; Pred. No. 3.8e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggcacaaacagccta 17
    |||||
DB 44 GCTGCAAAACAGCCCA 28

RESULT 12
5221624-15/c
; Patent No. 5221624
; APPLICANT: BLAIR, LINDLEY C.; KODURI, JAR-HOW; WEICKMANN,
```

```
; JOACHIM J.
; TITLE OF INVENTION: DNA ENCODING (LYS46, ASP97, ASP113) AND
; (LYS46, ASP113, ASP137) THAUMATIN I
; NUMBER OF SEQUENCES: 31
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/407,416
; FILING DATE: 14-SEP-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 268,702
; FILING DATE: 08-NOV-1988
; SEQ ID NO:15:
; LENGTH: 57
5221624-15

Query Match          67.8%; Score 12.2; DB 7; Length 57;
Best Local Similarity 82.4%; Pred. No. 3.8e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcgacaaaaccagccta 17
   || |||||
Db 44 GCTGCAAAACCGCCAA 28

RESULT 13
5221624-17/c
; Patent No. 5221624
; APPLICANT: BLAIR, LINDLEY C.; KODURI, JAR-HOW; WEICKMANN,
; JOACHIM J.
; TITLE OF INVENTION: DNA ENCODING (LYS46, ASP97, ASP113) AND
; (LYS46, ASP113, ASP137) THAUMATIN I
; NUMBER OF SEQUENCES: 31
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/407,416
; FILING DATE: 14-SEP-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 268,702
; FILING DATE: 08-NOV-1988
; SEQ ID NO:17:
; LENGTH: 57
5221624-17

Query Match          67.8%; Score 12.2; DB 7; Length 57;
Best Local Similarity 82.4%; Pred. No. 3.8e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcgacaaaaccagccta 17
   || |||||
Db 44 GCTGCAAAACCGCCAA 28

RESULT 14
5221624-19/c
; Patent No. 5221624
; APPLICANT: BLAIR, LINDLEY C.; KODURI, JAR-HOW; WEICKMANN,
; JOACHIM J.
; TITLE OF INVENTION: DNA ENCODING (LYS46, ASP97, ASP113) AND
; (LYS46, ASP113, ASP137) THAUMATIN I
; NUMBER OF SEQUENCES: 31
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/407,416
; FILING DATE: 14-SEP-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 268,702
; FILING DATE: 08-NOV-1988
; SEQ ID NO:19:
; LENGTH: 57
5221624-19

Query Match          67.8%; Score 12.2; DB 7; Length 57;
Best Local Similarity 82.4%; Pred. No. 3.8e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcgacaaaaccagccta 17
   || |||||
Db 44 GCTGCAAAACCGCCAA 28
```

```
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcgacaaaaccagccta 17
   || |||||
Db 44 GCTGCAAAACCGCCAA 28

RESULT 15
5221624-23/c
; Patent No. 5221624
; APPLICANT: BLAIR, LINDLEY C.; KODURI, JAR-HOW; WEICKMANN,
; JOACHIM J.
; TITLE OF INVENTION: DNA ENCODING (LYS46, ASP97, ASP113) AND
; (LYS46, ASP113, ASP137) THAUMATIN I
; NUMBER OF SEQUENCES: 31
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/407,416
; FILING DATE: 14-SEP-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 268,702
; FILING DATE: 08-NOV-1988
; SEQ ID NO:23:
; LENGTH: 63
5221624-23

Query Match          67.8%; Score 12.2; DB 7; Length 63;
Best Local Similarity 82.4%; Pred. No. 3.9e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

Search completed: June 4, 2000, 16:09:43  
Job time: 28062 sec



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:24:10 ; Search time 322.35 Seconds  
(without alignments)  
13.971 Million cell updates/sec

Title: US-09-164-714-13

Perfect score: 18  
Sequence: 1 ggtgtgtgtgtgtgttt 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 431286

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	13.8	76.7	36	1	055562	Sequence of synthe
2	13.8	76.7	36	1	055561	Sequence of contric
3	13.8	76.7	36	1	055569	Sequence of synthe
4	13.8	76.7	36	1	055574	Sequence of synthe
5	13.8	76.7	36	1	067225	Triple helix-form
6	13.8	76.7	36	1	067229	Triple helix-form
7	13.8	76.7	36	1	067232	Triple helix-form
8	13.8	76.7	36	1	079205	Guanosine rich oli
9	13.8	76.7	36	1	079207	Guanosine rich oli
10	13.8	76.7	36	1	T51623	Viral integrase in
11	13.8	76.7	36	1	T51625	Viral integrase in
12	13.8	76.7	42	1	X27667	l(trp) internal repl
13	13.4	74.4	27	1	V61397	DNA template 5'-en
14	13.4	74.4	27	1	V07668	Primer Mu-Bio used
15	13.4	74.4	27	1	X32517	Sequence of Mu-Bio
16	13.4	74.4	77	1	T20428	Human gene signatu
17	13.2	73.3	20	1	T97948	G-quartet oligonuc
18	13.2	73.3	24	1	N80521	Synthetic oligonuc
19	13.2	73.3	24	1	079216	Guanosine rich oli
20	13.2	73.3	35	1	T51634	Viral integrase in
21	13.2	73.3	35	1	033260	Triplex forming ol
22	13.2	73.3	41	1	T27080	S.hyicus lipase ge
23	13.2	73.3	41	1	V50702	Brassica sp. polym
24	13.2	73.3	45	1	069019	E. coli tryptophan
25	13.2	73.3	45	1	069017	E. coli tryptophan
26	13.2	73.3	46	1	094193	Transmembrane olig
27	13.2	73.3	67	1	V13178	Human cyclin E gene
28	13.2	73.3	72	1	X25976	Human IL-5 gene pr
29	13.2	73.3	78	1	X27805	CRCA-1 coding sequ
30	12.8	71.1	18	1	X09873	Human biallelic po
31	12.8	71.1	22	1	X32038	MUHI gene specific
32	12.8	71.1	30	1	V02229	Thyrotropin recept
33	12.8	71.1	35	1	004993	Sequence binding t
34	12.8	71.1	35	1	Q36271	APP5par, targeted

35	12.8	71.1	35	1	036272	APP5ant1, targete
36	12.8	71.1	38	1	005015	Sequence binding t
37	12.8	71.1	38	1	036337	HSVORLpar, target
38	12.8	71.1	38	1	036338	HSVORLant1, target
39	12.8	71.1	53	1	X23820	Bacillus sp. GUS 9
40	12.8	71.1	66	1	V77825	Staphylococcus aur
41	12.8	71.1	76	1	003555	DNA sequence encod
42	12.8	71.1	76	1	N92819	Modified type of s
43	12.8	71.1	77	1	003553	DNA sequence of 1m
44	12.8	71.1	77	1	N92360	Gene for modified
45	12.8	71.1	79	1	N81283	Sequence encoding

## ALIGNMENTS

RESULT	ID	Score	Match	Length	DB ID	Description
055562	1	13.8	76.7	36	1	055562
AC	055562	standard	DNA	36	BP.	
DT	17-AUG-1994	(first entry)				
DE	Sequence of synthetic triplex forming oligo (TFO) Seq ID no. 2.					
KW	Triplex forming oligonucleotide; TFO; sequence specific binding; ss.					
OS	Synthetic.					
FH	Key					
FT	misc_feature	36				
FT	/*tag= a					
PN	W09404550-A.					
PD	03-MAR-1994.					
PF	17-AUG-1993; U07743.					
PR	21-AUG-1992; US-934065.					
PR	23-APR-1993; US-053040.					
PA	(BAYU ) BAYLOR COLLEGE MEDICINE.					
PI	(TRIP-) TRIPLEX PHARM CORP.					
PI	Hogan ME, Jayaraman K, Vu H, Zengdegi J;					
DR	WPI: 94-083097/10					
PT	New method of binding synthetic triplex forming oligonucleotide -					
PT	in which the nucleotide is modified with a lipophilic cpd. is					
PT	useful in treatment of cell proliferative states and viral					
PS	infections					
PS	Claim 8: Page 24; 86pp: English.					
CC	Triplex forming oligos (TFOs) bind to DNA in a site selective manner.					
CC	The biological effect of a TFO is potentiated by modification with					
CC	lipophilic cpds., selected from cholesterol, vitamin E and 1,2-di-O-					
CC	hexadecyl-3-glycerol. TFO-linker-cholesterol is used for the					
CC	treatment of cell proliferative states (breast, lung and cervical					
CC	cancers) and in infections by viruses (Herpes simplex virus type 2					
CC	and human immunodeficiency virus (HIV)).					
CC	Sequence 36 BP; 0 A; 0 C; 24 G; 12 T;					
SQ						
Query Match		76.7%	Score 13.8;	DB 1;	Length 36;	
Best Local Similarity		88.2%	Pred. No. 4.4e+02;			
Matches 15;	Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;	
QY	1 ggtgtgtgtgtgtgtt 17					
Db	9 GGTGTGTGTGTGTGT 25					
RESULT	2					
055561						
ID	055561	standard	DNA	36	BP.	
AC	055561					
DT	17-AUG-1994	(first entry)				
DE	Sequence of control for synthetic triplex forming oligo (TFO)					
DE	Seq ID no. 2.					
KW	Triplex forming oligonucleotide; TFO; sequence specific binding; ss.					
OS	Synthetic.					
FH	Key					
FT	misc_feature	36				
FT	/*tag= a					

FT /label= G-NH2.  
 PN WO9404550-A.  
 PD 03-MAR-1994.  
 PF 17-AUG-1993; U07743.  
 PR 21-AUG-1992; US-934065.  
 PR 23-APR-1993; US-053040.  
 PA (BAYU ) BAYLOR COLLEGE MEDICINE.  
 PI (TRIP-) TRIPLEX PHARM CORP.  
 DR Hogan ME, Jayaraman K, Vu H, Zengdegi J;  
 PI: 94-083097/10.  
 PT New method of binding synthetic triplex forming oligonucleotide -  
 PT in which the nucleotide is modified with a lipophilic cpd. is  
 PT useful in treatment of cell proliferative states and viral  
 PT infections  
 PS Example; Page 24; 86pp; English.  
 CC Triplex forming oligos (TFOs) bind to DNA in a site selective manner.  
 CC The biological effect of a TFO is potentiated by modification with  
 CC lipophilic cpds.. selected from cholesterol, vitamin E and 1,2-di-O-  
 CC hexadecyl-3-glyceryl. TFO-linker-cholesterol is used for the  
 CC treatment of cell proliferative states (breast, lung and cervical  
 CC cancers) and in infections by viruses (Herpes simplex virus type 2  
 CC and human immunodeficiency virus (HIV)).  
 SQ Sequence 36 BP; 0 A; 0 C; 24 G; 12 T;

Query Match 76.7%; Score 13.8; DB 1; Length 36;  
 Best Local Similarity 88.2%; Pred. No. 4.4e+02;  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ggtgtgtgtgtgtgtgt 17  
 |||||  
 DB 9 GGTGTTGTCGTGCTTT 25

RESULT 3  
 Q55569  
 ID Q55569 standard; DNA; 36 BP.  
 AC Q55569;  
 DE 17-AUG-1994 (first entry)  
 DE Sequence of synthetic triplex forming oligo (TFO) B-106-85,  
 DE anti-HSV 2 TFO  
 KM Triplex forming oligonucleotide; TFO; sequence specific binding; ss.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_feature 36  
 FT /\*tag= a  
 FT /label= G-cholesterol  
 PN WO9404550-A.  
 PD 03-MAR-1994.  
 PF 17-AUG-1993; U07743.  
 PR 21-AUG-1992; US-934065.  
 PR 23-APR-1993; US-053040.  
 PA (BAYU ) BAYLOR COLLEGE MEDICINE.  
 PI (TRIP-) TRIPLEX PHARM CORP.  
 DR Hogan ME, Jayaraman K, Vu H, Zengdegi J;  
 PI: 94-083097/10.  
 PT New method of binding synthetic triplex forming oligonucleotide -  
 PT in which the nucleotide is modified with a lipophilic cpd. is  
 PT useful in treatment of cell proliferative states and viral  
 PT infections  
 PS Example; Page 40; 86pp; English.  
 CC Triplex forming oligos (TFOs) bind to DNA in a site selective manner.  
 CC The biological effect of a TFO is potentiated by modification with  
 CC lipophilic cpds.. selected from cholesterol, vitamin E and 1,2-di-O-  
 CC hexadecyl-3-glyceryl. TFO-linker-cholesterol is used for the  
 CC treatment of cell proliferative states (breast, lung and cervical  
 CC cancers) and in infections by viruses (Herpes simplex virus type 2  
 CC and human immunodeficiency virus (HIV)).  
 SQ Sequence 36 BP; 0 A; 0 C; 24 G; 12 T;

Query Match 76.7%; Score 13.8; DB 1; Length 36;  
 Best Local Similarity 88.2%; Pred. No. 4.4e+02;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 OY 1 ggtgtgtgtgtgtgt 17  
 |||||  
 DB 9 GGTGTTGTCGTGCTTT 25

RESULT 4  
 Q55574  
 ID Q55574 standard; DNA; 36 BP.  
 AC Q55574;  
 DE 17-AUG-1994 (first entry)  
 DE Sequence of synthetic triplex forming oligo (TFO)  
 DE B-106-89, anti-HIV2 TFO  
 KM Triplex forming oligonucleotide; TFO; sequence specific binding; ss.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_feature 1  
 FT /\*tag= a  
 FT /label= Cholesterol-G  
 FT misc\_feature 36  
 FT /\*tag= b  
 FT /label= G-propanolamine  
 PN WO9404550-A.  
 PD 03-MAR-1994.  
 PF 17-AUG-1993; U07743.  
 PR 21-AUG-1992; US-934065.  
 PR 23-APR-1993; US-053040.  
 PA (BAYU ) BAYLOR COLLEGE MEDICINE.  
 PI (TRIP-) TRIPLEX PHARM CORP.  
 DR Hogan ME, Jayaraman K, Vu H, Zengdegi J;  
 PI: 94-083097/10.  
 PT New method of binding synthetic triplex forming oligonucleotide -  
 PT in which the nucleotide is modified with a lipophilic cpd. is  
 PT useful in treatment of cell proliferative states and viral  
 PT infections  
 PS Example; Page 41; 86pp; English.  
 CC Triplex forming oligos (TFOs) bind to DNA in a site selective manner.  
 CC The biological effect of a TFO is potentiated by modification with  
 CC lipophilic cpds.. selected from cholesterol, vitamin E and 1,2-di-O-  
 CC hexadecyl-3-glyceryl. TFO-linker-cholesterol is used for the  
 CC treatment of cell proliferative states (breast, lung and cervical  
 CC cancers) and in infections by viruses (Herpes simplex virus type 2  
 CC and human immunodeficiency virus (HIV)). Q55573-75 were synthesised  
 CC using the cholesterolyl phosphoramidite, synthesis of  
 CC 2-cyanoethyl-N,N'-diisopropyl-1-O-(4,4'-dimethoxytrityl)-2'-N-  
 CC ((N'-cholesterylloxycarbonyl)glycidylglycidyl)aminopropaneglycyl  
 CC phosphoramidite.  
 SQ Sequence 36 BP; 0 A; 0 C; 24 G; 12 T;

Query Match 76.7%; Score 13.8; DB 1; Length 36;  
 Best Local Similarity 88.2%; Pred. No. 4.4e+02;  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 OY 1 ggtgtgtgtgtgtgt 17  
 |||||  
 DB 9 GGTGTTGTCGTGCTTT 25

RESULT 5  
 Q67225  
 ID Q67225 standard; DNA; 36 BP.  
 AC Q67225;  
 DE 22-SEP-1994 (first entry)  
 DE Triplex helix-forming oligonucleotide.  
 DE Triple helix; oligonucleotide; TFO; lipophilic; cholesterol; Vitamin E;  
 KM glyceryl; anticancer; antiviral; herpes simplex; HIV; human;  
 KM immunodeficiency virus.  
 OS Synthetic.  
 PN WO9404550-A.  
 PD 03-MAR-1994.  
 PF 17-AUG-1993; U07743.

PR 21-AUG-1992; US-934065.  
PR 23-APR-1993; US-053040.  
PA (BAYU ) BAYLOR COLLEGE MEDICINE.  
PA (TRIP-) TRIPLEX PHARM CORP.  
PI Hogan ME, Jayaraman K, Vu H, Zengdegi J;  
DR WPI: 94-083097/10.  
PT New method of binding synthetic triplex forming oligonucleotide -  
PT in which the nucleotide is modified with a lipophilic cpd; is  
PT useful in treatment of cell proliferative states and viral  
PT infections  
PS Claim 8; Page 24; 86pp; English.  
CC The invention relates to a method of enhancing sequence specific  
CC binding of a synthetic triplex-forming oligonucleotide (TFO)  
CC involving the step of contacting the TFO with a cell. The TFO comprises  
CC a nucleotide sequence of at least 20 nucleotides including a G and a T,  
CC is chemically modified with a lipophilic compound and is capable of  
CC binding to a DNA duplex target to form a triplex helix. The lipophilic  
CC compound is preferably cholesterol, vitamin E or 1,2-di-O-hexadecyl  
CC 3-glycerol and is joined to the nucleotide via a linker.  
CC The TFO is useful medically for treating cell proliferative  
CC states such as breast, lung and cervical cancers, and for treating  
CC viral infections such as caused by Herpes simplex type 2 or HIV. The  
CC biological effect of the TFO is potentiated by modification with the  
CC lipophilic compound.  
CC The present sequence is one of 9 specific nucleotide sequences disclosed  
CC for use in forming the TFO (067224 - 067232). The TFO is formed by  
CC attaching cholesterol to the 3' end. 24 G; 12 T;  
SQ Sequence 36 BP; 0 A; 0 C; 24 G; 12 T;

Query Match 76.7%; Score 13.8; DB 1; Length 36;  
Best Local Similarity 88.2%; Pred. No. 4.4e+02;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ggtgtgtgtgtgtgt 17  
|||||||  
Db 9 GGTGTGTGTGTGTGT 25

RESULT 6  
067229  
ID 067229 standard; DNA; 36 BP.  
AC 067229;  
DT 22-SEP-1994 (first entry)  
DE Triple helix-forming oligonucleotide.  
CC Triple helix; oligonucleotide; TFO; lipophilic; cholesterol; Vitamin E;  
KW glyceryl; anticancer; antiviral; herpes simplex; HIV; human;  
KM immunodeficiency virus.  
OS Synthetic.  
PN WO9404550-A.  
PD 03-MAR-1994.  
PF 17-AUG-1993; U07743.  
PR 21-AUG-1992; US-934065.  
PR 23-APR-1993; US-053040.  
PA (BAYU ) BAYLOR COLLEGE MEDICINE.  
PA (TRIP-) TRIPLEX PHARM CORP.  
PI Hogan ME, Jayaraman K, Vu H, Zengdegi J;  
DR WPI: 94-083097/10.  
PT New method of binding synthetic triplex forming oligonucleotide -  
PT in which the nucleotide is modified with a lipophilic cpd; is  
PT useful in treatment of cell proliferative states and viral  
PT infections  
PS Claim 25; Page 40; 86pp; English.  
CC The invention relates to a method of enhancing sequence specific  
CC binding of a synthetic triplex-forming oligonucleotide (TFO)  
CC involving the step of contacting the TFO with a cell. The TFO comprises  
CC a nucleotide sequence of at least 20 nucleotides including a G and a T,  
CC is chemically modified with a lipophilic compound and is capable of  
CC binding to a DNA duplex target to form a triplex helix. The lipophilic  
CC compound is preferably cholesterol, vitamin E or 1,2-di-O-hexadecyl  
CC 3-glycerol and is joined to the nucleotide via a linker.  
CC The TFO is useful medically for treating cell proliferative  
CC states such as breast, lung and cervical cancers, and for treating

CC viral infections such as caused by Herpes simplex type 2 or HIV. The  
CC biological effect of the TFO is potentiated by modification with the  
CC lipophilic compound.  
CC The present sequence is one of 9 specific nucleotide sequences disclosed  
CC for use in forming the TFO (067224 - 067232). The TFO is formed by  
CC attaching cholesterol to the 3' end. This TFO is active against  
CC HSV-2.  
SQ Sequence 36 BP; 0 A; 0 C; 24 G; 12 T;

Query Match 76.7%; Score 13.8; DB 1; Length 36;  
Best Local Similarity 88.2%; Pred. No. 4.4e+02;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ggtgtgtgtgtgtgt 17  
|||||||  
Db 9 GGTGTGTGTGTGTGT 25

RESULT 7  
067232  
ID 067232 standard; DNA; 36 BP.  
AC 067232;  
DT 22-SEP-1994 (first entry)  
DE Triple helix-forming oligonucleotide.  
CC Triple helix; oligonucleotide; TFO; lipophilic; cholesterol; Vitamin E;  
KW glyceryl; anticancer; antiviral; herpes simplex; HIV; human;  
KM immunodeficiency virus.  
OS Synthetic.  
PN WO9404550-A.  
PD 03-MAR-1994.  
PF 17-AUG-1993; U07743.  
PR 21-AUG-1992; US-934065.  
PR 23-APR-1993; US-053040.  
PA (BAYU ) BAYLOR COLLEGE MEDICINE.  
PA (TRIP-) TRIPLEX PHARM CORP.  
PI Hogan ME, Jayaraman K, Vu H, Zengdegi J;  
DR WPI: 94-083097/10.  
PT New method of binding synthetic triplex forming oligonucleotide -  
PT in which the nucleotide is modified with a lipophilic cpd; is  
PT useful in treatment of cell proliferative states and viral  
PT infections  
PS Claim 25; Page 41; 86pp; English.  
CC The invention relates to a method of enhancing sequence specific  
CC binding of a synthetic triplex-forming oligonucleotide (TFO)  
CC involving the step of contacting the TFO with a cell. The TFO comprises  
CC a nucleotide sequence of at least 20 nucleotides including a G and a T,  
CC is chemically modified with a lipophilic compound and is capable of  
CC binding to a DNA duplex target to form a triplex helix. The lipophilic  
CC compound is preferably cholesterol, vitamin E or 1,2-di-O-hexadecyl  
CC 3-glycerol and is joined to the nucleotide via a linker.  
CC The TFO is useful medically for treating cell proliferative  
CC states such as breast, lung and cervical cancers, and for treating  
CC viral infections such as caused by Herpes simplex type 2 or HIV. The  
CC biological effect of the TFO is potentiated by modification with the  
CC lipophilic compound.  
CC The present sequence is one of 9 specific nucleotide sequences disclosed  
CC for use in forming the TFO (067224 - 067232). The TFO is formed by  
CC attaching cholesterol to the 5' end and propanolamine to the 3' end.  
CC This TFO is active against HSV-2.  
SQ Sequence 36 BP; 0 A; 0 C; 24 G; 12 T;

Query Match 76.7%; Score 13.8; DB 1; Length 36;  
Best Local Similarity 88.2%; Pred. No. 4.4e+02;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ggtgtgtgtgtgtgt 17  
|||||||  
Db 9 GGTGTGTGTGTGTGT 25

RESULT 8

```
Q79205
ID Q79205 standard; DNA; 36 BP.
AC Q79205;
DE 17-JUL-1995 (first entry)
KW Guanostine rich oligonucleotide used to treat viral infection.
KW Guanostine; tetrad; inhibition; replication; virus; treatment;
KW therapy; infection; herpes simplex virus; human papilloma virus;
KW Epstein-Barr virus; HIV; adenovirus; respiratory syncytial virus;
KW hepatitis B virus; human cytomegalovirus; ss.
OS Synthetic.
FH Key
FH Key Location/Qualifiers
FT misc_feature 36
FT /*tag= a
FT /mod_base=
FT /note= "Propanolamine group attached to this base."
PN MO9425037-A.
PD 10-NOV-1994.
PE 25-APR-1994; US-053027.
PR 23-APR-1993; US-053027.
PR 28-OCT-1993; US-145704.
PA (BAYU ) BAYLOR COLLEGE MEDICINE.
PA (TRIP-) TRIPLEX PHARM CORP.
PI Fennewald S, Hogan ME, Ojwang JO, Rando RF, Zendegui JG;
PI WPI: 94-357890/44.
PT Oligo-nucleotide(s) rich in guanosine which form guanosine
PT tetrads - used to treat viral infections, e.g. herpes-virus and
PT HIV
PS Claim 41; Page 46; 101pp; English.
CC The oligonucleotides (See Q79201-52) can be used to treat viral
CC infections. The oligonucleotides inhibit viral replication by
CC forming guanosine tetrads which form a stabilised 3D structure.
CC Preferred oligonucleotides contain at least 2 runs of at least 2
CC guanosine bases and may be capped at the 3' terminus with a modifier
CC selected from polyamine, poly-L-lysine, cholesterol and
CC propanolamine. They may also have a modified phosphodiester linkage
CC or be modified to contain a phosphorothioate linkage. They are used
CC to treat infections with viruses such as herpes simplex virus, human
CC papilloma virus, Epstein-Barr virus, HIV, adenovirus, respiratory
CC syncytial virus, hepatitis B virus or human cytomegalovirus.
SQ Sequence 36 BP; 0 A; 0 C; 24 G; 12 T;

Query Match 76.7%; Score 13.8; DB 1; Length 36;
Best Local Similarity 88.2%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggtgtgtgtgtgtgctt 17
   |||||
Db 9 GGTGTGTGTGTGTGTTT 25

RESULT 9
Q79207
ID Q79207 standard; DNA; 36 BP.
AC Q79207;
DE 17-JUL-1995 (first entry)
KW Guanostine rich oligonucleotide used to treat viral infection.
KW Guanostine; tetrad; inhibition; replication; virus; treatment;
KW therapy; infection; herpes simplex virus; human papilloma virus;
KW Epstein-Barr virus; HIV; adenovirus; respiratory syncytial virus;
KW hepatitis B virus; human cytomegalovirus; ss.
OS Synthetic.
FH Key
FH Key Location/Qualifiers
FT misc_feature 36
FT /*tag= a
FT /mod_base=
FT /note= "Propanolamine group attached to this base."
PN MO9425037-A.
PD 10-NOV-1994.
PE 25-APR-1994; US-053027.
PR 23-APR-1993; US-053027.
PR 28-OCT-1993; US-145704.
PA (BAYU ) BAYLOR COLLEGE MEDICINE.
PA (TRIP-) TRIPLEX PHARM CORP.
PI Fennewald S, Hogan ME, Ojwang JO, Rando RF, Zendegui JG;
PI WPI: 94-357890/44.
PT Oligo-nucleotide(s) rich in guanosine which form guanosine
PT tetrads - used to treat viral infections, e.g. herpes-virus and
PT HIV
PS Claim 41; Page 46; 101pp; English.
CC The oligonucleotides (See Q79201-52) can be used to treat viral
CC infections. The oligonucleotides inhibit viral replication by
CC forming guanosine tetrads which form a stabilised 3D structure.
CC Preferred oligonucleotides contain at least 2 runs of at least 2
CC guanosine bases and may be capped at the 3' terminus with a modifier
CC selected from polyamine, poly-L-lysine, cholesterol and
CC propanolamine. They may also have a modified phosphodiester linkage
CC or be modified to contain a phosphorothioate linkage. They are used
CC to treat infections with viruses such as herpes simplex virus, human
CC papilloma virus, Epstein-Barr virus, HIV, adenovirus, respiratory
CC syncytial virus, hepatitis B virus or human cytomegalovirus.
SQ Sequence 36 BP; 0 A; 0 C; 24 G; 12 T;
```

```
PR 23-APR-1993; US-053027.
PR 28-OCT-1993; US-145704.
PA (BAYU ) BAYLOR COLLEGE MEDICINE.
PA (TRIP-) TRIPLEX PHARM CORP.
PI Fennewald S, Hogan ME, Ojwang JO, Rando RF, Zendegui JG;
PI WPI: 94-357890/44.
PT Oligo-nucleotide(s) rich in guanosine which form guanosine
PT tetrads - used to treat viral infections, e.g. herpes-virus and
PT HIV
PS Claim 41; Page 47; 101pp; English.
CC The oligonucleotides (See Q79201-52) can be used to treat viral
CC infections. The oligonucleotides inhibit viral replication by
CC forming guanosine tetrads which form a stabilised 3D structure.
CC Preferred oligonucleotides contain at least 2 runs of at least 2
CC guanosine bases and may be capped at the 3' terminus with a modifier
CC selected from polyamine, poly-L-lysine, cholesterol and
CC propanolamine. They may also have a modified phosphodiester linkage
CC or be modified to contain a phosphorothioate linkage. They are used
CC to treat infections with viruses such as herpes simplex virus, human
CC papilloma virus, Epstein-Barr virus, HIV, adenovirus, respiratory
CC syncytial virus, hepatitis B virus or human cytomegalovirus.
SQ Sequence 36 BP; 0 A; 0 C; 24 G; 12 T;

Query Match 76.7%; Score 13.8; DB 1; Length 36;
Best Local Similarity 88.2%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggtgtgtgtgtgtgctt 17
   |||||
Db 9 GGTGTGTGTGTGTGTTT 25

RESULT 10
T51623
ID T51623 standard; DNA; 36 BP.
AC T51623;
DE 11-NOV-1997 (first entry)
KW Viral integrase inhibiting oligonucleotide.
KW Human immunodeficiency virus; HIV; Epstein Barr virus; EBV;
KW herpes simplex virus; HSV; human papilloma virus; HPV; adenovirus;
KW respiratory syncytial virus; RSV; cytomegalovirus; CMV; hepatitis B;
KW integrase inhibition; guanosine tetrad; ss.
OS Synthetic.
FH Key
FH Key Location/Qualifiers
FT misc_feature 36
FT /*tag= a
FT /mod_base=
FT /note= "Propanolamine group attached to this base."
PN MO9703997-A1.
PD 06-FEB-1997.
PE 17-JUL-1996; U11786.
PE 23-APR-1996; US-016271.
PR 19-JUL-1995; US-001505.
PR 23-OCT-1995; US-535168.
PR 19-MAR-1996; US-013688.
PR 25-MAR-1996; US-014007.
PR 17-APR-1996; US-015714.
PA (ARON-) ARONEX PHARM INC.
PA Fennewald S, Hogan ME, Mazumder A, Ojwang JO, Pommler Y;
PI Rando RF, Zendegui JG;
PI WPI: 97-132569/12.
PT Oligo-nucleotide(s) capable of forming guanosine tetrads - inhibit
PT viral enzyme responsible for integrating viral nucleic acid into the
PT host genome
PS Claim 3; Page 142; 245pp; English.
CC T51619-T51698 are oligonucleotides used to inhibit the production
CC of viruses within a host cell. The oligonucleotides may form guanosine
CC tetrads (structures formed of eight hydrogen bonds by coordination of
CC the four oxygen atoms of guanine with alkali cations believed to bind
CC to the centre of a quadruplex, and by strong stacking interactions) and
CC are used to prevent the integration of viral nucleic acid into a host
CC genome. The oligonucleotides inhibit functioning of the integrase enzyme
CC and hence prevent viral infection. Viral infections that may be treated
CC include human immunodeficiency virus (HIV), Epstein Barr virus (EBV),
CC herpes simplex virus (HSV), human papilloma virus (HPV), adenovirus,
CC respiratory syncytial virus (RSV), cytomegalovirus (CMV) and hepatitis
CC B virus (HBV), especially HIV-1 infection.
```

SQ Sequence 36 BP: 0 A; 0 C; 24 G; 12 T;  
 Query Match 76.7%; Score 13.8; DB 1; Length 36;  
 Best Local Similarity 88.2%; Pred. No. 4.4e+02;  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 ggtgtgtgtgtgtgtgt 17  
 ||| ||| ||| ||| ||| |||  
 Db 9 GGTGTGTGTGTGTGT 25  
 RESULT 11  
 T51625  
 ID T51625 standard; DNA: 36 BP.  
 AC T51625;  
 DE 11-NOV-1997 (first entry)  
 KW Human immunodeficiency virus; HIV; Epstein Barr virus; EBV;  
 KW herpes simplex virus; HSV; human papilloma virus; HPV; adenovirus;  
 KW respiratory syncytial virus; RSV; cytomegalovirus; CMV; hepatitis B;  
 KW integrase inhibition; guanosine tetrad; ss.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_difference 36  
 FT /\*tag- a  
 FT /note= "phosphorothioate backbone"  
 PN WO9703997-A1.  
 PD 06-FEB-1997.  
 PE 17-JUL-1996; U11786.  
 PR 23-APR-1996; US-016271.  
 PR 19-JUL-1995; US-001505.  
 PR 23-OCT-1995; US-535168.  
 PR 19-MAR-1996; US-013688.  
 PR 25-MAR-1996; US-014007.  
 PR 17-APR-1996; US-015714.  
 PA (ARON-) ARONEX PHARM INC.  
 PI Fennewald S, Hogan ME, Mazumder A, O'wang JO, Pommer Y;  
 PI Rando RF, Zendejani JG;  
 PI WPI: 97-132569/12.  
 PT Oligo:nucleotide(s) capable of forming guanosine tetrads - inhibit  
 PT viral enzyme responsible for integrating viral nucleic acid into the  
 PT host genome  
 PS Claim 3; Page 143; 245pp; English.  
 CC T51619-T51698 are oligonucleotides used to inhibit the production  
 CC of viruses within a host cell. The oligonucleotides may form guanosine  
 CC tetrads (structures formed of eight hydrogen bonds by coordination of  
 CC the four oxygen atoms of guanine with alkali cations believed to bind  
 CC to the centre of a quadruplex, and by strong stacking interactions) and  
 CC are used to prevent the integration of viral nucleic acid into a host  
 CC genome. The oligonucleotides inhibit functioning of the integrase enzyme  
 CC and hence prevent viral infection. Viral infections that may be treated  
 CC include human immunodeficiency virus (HIV), Epstein Barr virus (EBV),  
 CC herpes simplex virus (HSV), human papilloma virus (HPV), adenovirus,  
 CC respiratory syncytial virus (RSV), cytomegalovirus (CMV) and hepatitis  
 CC B virus (HBV), especially HIV-1 infection.  
 CC Sequence 36 BP; 0 A; 0 C; 24 G; 12 T;  
 SQ  
 Query Match 76.7%; Score 13.8; DB 1; Length 36;  
 Best Local Similarity 88.2%; Pred. No. 4.4e+02;  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 ggtgtgtgtgtgtgtgt 17  
 ||| ||| ||| ||| ||| |||  
 Db 9 GGTGTGTGTGTGTGT 25  
 RESULT 12  
 X27667/C  
 ID X27667 standard; DNA: 42 BP.  
 AC X27667;  
 DE 01-JUN-1999 (first entry)

DE (Trp) internal repeat oligo used in synthetic HRP gene modules.  
 KW Synthetic gene; plant; gum; hydroxyproline-rich glycoprotein; HRP;  
 KW repetitive proline-rich protein; RRP; arabinogalactan protein; AGP;  
 KW glycopeptide; ss.  
 OS Synthetic.  
 PN WO9903978-A1.  
 PD 28-JAN-1999.  
 PE 21-JUL-1998; U15083.  
 PR 20-JUL-1998; US-897556.  
 PR 21-JUL-1997; US-897556.  
 PA (UYOH-) UNIV OHIO.  
 PI Kieleszewski MJ;  
 PI WPI: 99-132225/11.  
 PT Novel synthetic gene designed from repetitive peptide sequences - of  
 PT hydroxyproline-rich glycoprotein  
 PS Example 4; Page 38; 72pp; English.  
 CC The invention relates to novel synthetic genes for plant gums. A new  
 CC approach is described to the production of hydroxyproline-rich  
 CC glycoproteins (HRGPs), repetitive proline-rich proteins (RPRPs) and  
 CC arabinogalactan proteins (AGPs). Synthetic genes comprising a nucleic  
 CC acid encoding the peptide (Y01267) can be engineered for the production  
 CC of repetitive glycopeptide modules in cells. The invention provided a new  
 CC approach to the problem of producing plant gums that is not dependent on  
 CC environmental factors and greatly simplifies the production of a variety  
 CC of naturally occurring gums as well as designer gums.  
 CC Sequence 42 BP; 13 A; 24 C; 1 G; 4 T;  
 SQ  
 Query Match 76.7%; Score 13.8; DB 1; Length 42;  
 Best Local Similarity 88.2%; Pred. No. 4.4e+02;  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 ggtgtgtgtgtgtgtgt 17  
 ||| ||| ||| ||| ||| |||  
 Db 35 GGTGTGTGTGTGTGT 19  
 RESULT 13  
 V61397  
 ID V61397 standard; DNA: 24 BP.  
 AC V61397;  
 DE 17-DEC-1998 (first entry)  
 KW DNA template 5'-end region from WO9844151 Example 1.  
 KW Nucleic acid amplification; immobilised primer; sequencing;  
 KW screening; diagnosis; in situ nucleic acid synthesis; fingerprinting;  
 KW gene expression; ss.  
 OS Synthetic.  
 PN WO9844151-A1.  
 PD 08-OCT-1998.  
 PE 01-APR-1998; G00961.  
 PR 23-JUN-1997; GB-013238.  
 PR 01-APR-1997; GB-006528.  
 PR 01-APR-1997; GB-006529.  
 PR 23-JUN-1997; GB-013236.  
 PA (GLAXO) GLAXO GROUP LTD.  
 PI Farinelli L, Kawashima E, Mayer P;  
 PI WPI: 98-568282/48.  
 PT New nucleic acid amplification - by extension of immobilised  
 PT primers annealed to target, separation of strands, annealing  
 PT extended primer to second primer and repeating extension  
 PS Example 1; Page 23; 108pp; English.  
 CC A method has been developed for nucleic acid (NA) amplification. The  
 CC method comprises: (a) annealing single-stranded (ss) target NA to one of  
 CC many primers (P), immobilised but having one end free for extension, and  
 CC extending P using the target as template, producing an immobilised  
 CC extension product (EP); (b) separating EP and target NA; (c) annealing  
 CC EP to another P and extending this, using EP as template, to give a  
 CC second extension product (EP2); and (d) optionally separating EP and EP2.  
 CC The method is used to produce NA for sequencing, diagnosis and screening,  
 CC as supports for other materials, for generating free NA (particularly in  
 CC situ RNA synthesis), monitoring gene expression, identifying NA encoding  
 CC rarely expressed gene products and identifying heterozygotes and in NA  
 CC fingerprinting. A very high density of DNA-bearing areas can be randomly

CC arrayed and amplified on the support. The present sequence represents the  
CC 5'-end of a DNA template used in an example from the present invention.  
SQ Sequence 24 Bp; 0 A; 0 C; 12 G; 12 T;

Query Match  
Best Local Similarity 93.3%; Pred. No. 6.3e+02; Length 24;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 tctgtgtgtgtgctt 17  
|||  
Db 1 TGTGTGCTGTGCTT 15

RESULT 14  
V07668/c  
ID V07668 standard; DNA; 27 Bp.  
AC V07668:  
DT 21-DEC-1998 (first entry)  
DE Primer Mu-Bio used to amplify Mu-tagged sequence.  
KW Bxl gene; DIMBOA; 2,4-dihydroxy-7-methoxy-1,4 benzoxazin-3-one;  
KW maize; benzoxazinone; insect resistance; herbicide resistance;  
KW crop protection; transgenic plant; ss.  
OS Synthetic.  
PN W09840505-A1.  
PD 17-SEP-1998.  
PE 13-MAR-1998; 005078.  
PR 13-MAR-1997; US-040513.  
PA (DEKA-) DEKALB GENETICS CORP.  
PI Chomet P, Frey M, Gierl A;  
DR WPI; 98-520823/44.  
PT New isolated maize Bxl gene - used, e.g. for producing plants with  
PT increased resistance to pests, pathogens or chemicals  
PS Example 10; Page 94; 137pp; English.  
CC Primers Mu-Bio, Msel Sel/A(GCT), BfaI Sel/A(GCT) and Mu Sel (see  
CC V07668-71) were used to amplify Mu-tagged sequences in a Mu-tagged  
CC bxl allele of maize. The bxl gene (see V07665) is involved in  
CC the biosynthesis of the benzoxazinone DIMBOA. The gene provides a  
CC tool for the production of transgenic plants with enhanced  
CC expression profiles of benzoxazinone biosynthesis, and hence  
CC resistance to pest infestation, herbicide damage and disease.  
SQ Sequence 27 Bp; 8 A; 10 C; 5 G; 3 T;

Query Match 74.4%; Score 13.4; DB 1; Length 27;  
Best Local Similarity 82.4%; Pred. No. 6.4e+02;  
Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 ggtgtgtgtgtgctt 17  
|||  
Db 21 GGCGWTGGCGTTGGCTT 5

RESULT 15  
X32517/c  
ID X32517 standard; DNA; 27 Bp.  
AC X32517:  
DT 23-JUN-1999 (first entry)  
DE Sequence of Mu-Bio.  
KW Insertion; detection array; plant; screening; animal feed; bar gene;  
KW pharmaceutical; industrial product; T-DNA; primer; ss.  
OS Synthetic.  
PN W09914373-A1.  
PD 25-MAR-1999.  
PE 17-SEP-1998; U19369.  
PR 17-SEP-1997; US-932280.  
PA (UTYA ) UNITV YALE.  
PI Dellaporta SL;  
DR WPI; 99-229553/19.  
PT Screening for insertion genomic mutations  
PS Example 4; Page 43; 84pp; English.  
CC The invention relates to detection of an insertion event in a genome by  
CC screening at least one detection array comprising a DNA composition

CC enhanced for many insertion junctions. The method is used, particularly  
CC for plants: (1) to determine the function of particular genes and (2) to  
CC identify insertional mutations that confer a desirable property. The  
CC genes or mutations identified may then be used to generate new plants for  
CC uses as foods, animal feeds, e.g. as silage, or as sources of oil,  
CC starch, pharmaceuticals, industrial products etc. Unlike known methods,  
CC this process allows simultaneous screening of many unique insertional  
CC events, and sequences of single members of a population can be identified  
CC without individual analysis of each member of the population. The method  
CC is generally applicable to animals, lower eukaryotes, prokaryotes and  
CC particularly plants. Sequences X32516-522 represent primer and adapter  
CC sequences. It is indicated that these sequences are in their -3'  
SQ Sequence 27 Bp; 8 A; 10 C; 5 G; 3 T;

Query Match 74.4%; Score 13.4; DB 1; Length 27;  
Best Local Similarity 82.4%; Pred. No. 6.4e+02;  
Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 ggtgtgtgtgtgctt 17  
|||  
Db 21 GGCGWTGGCGTTGGCTT 5

Search completed: June 4, 2000, 16:24:11  
Job time: 28907 sec



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:09:43 ; Search time 244.64 Seconds  
(without alignments)  
9.564 Million cell updates/sec

Title: US-09-164-714-13

Perfect score: 18  
Sequence: 1 ggtgtgtgtgtgtgttt 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 375880

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/lna/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/lna/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/lna/5E.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/lna/5F.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/lna/5G.COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14.4	80.0	90	1	US-08-123-702-16 Sequence 16, Appl
2	13.8	76.7	36	1	US-08-145-704-5 Sequence 5, Appl
3	13.8	76.7	36	1	US-08-145-704-7 Sequence 7, Appl
4	13.8	76.7	36	6	PCT-US93-07743-7 Sequence 7, Appl
5	13.8	76.7	36	6	PCT-US93-07743-11 Sequence 11, Appl
6	13.8	76.7	36	6	PCT-US93-07743-16 Sequence 16, Appl
7	13.8	76.7	36	6	PCT-US96-11786-5 Sequence 5, Appl
8	13.8	76.7	36	6	PCT-US96-11786-7 Sequence 7, Appl
9	13.2	73.3	20	2	US-08-703-755A-4 Sequence 4, Appl
10	13.2	73.3	24	1	US-08-145-704-16 Sequence 16, Appl
11	13.2	73.3	24	6	PCT-US96-11786-16 Sequence 16, Appl
12	13.2	73.3	39	2	US-08-411-727-11 Sequence 11, Appl
13	13.2	73.3	45	1	US-08-197-770-6 Sequence 6, Appl
14	13.2	73.3	45	1	US-08-197-770-8 Sequence 8, Appl
15	13.2	73.3	46	4	US-08-495-695E-23 Sequence 23, Appl
16	13.2	73.3	46	6	PCT-US94-14436-23 Sequence 23, Appl
17	13.2	73.3	51	7	5474913-7 Patent No. 5474913
18	13.2	73.3	51	7	5474913-12 Patent No. 5474913
19	12.8	71.1	26	1	US-08-467-420A-43 Sequence 43, Appl
20	12.8	71.1	26	2	US-08-470-110A-43 Sequence 43, Appl
21	12.8	71.1	26	2	US-08-667-769A-43 Sequence 43, Appl
22	12.8	71.1	26	6	US-08-840-371-43 Sequence 43, Appl
23	12.8	71.1	26	6	PCT-US95-17082A-43 Sequence 43, Appl
24	12.8	71.1	37	6	PCT-US93-07743-2 Sequence 2, Appl
25	12.8	71.1	76	7	5182195-8 Patent No. 5182195
26	12.8	71.1	77	7	5182195-3 Patent No. 5182195
27	12.4	68.9	21	1	US-08-094-534-9 Sequence 9, Appl

28	12.4	68.9	21	4	US-08-581-543-9 Sequence 9, Appl
29	12.4	68.9	21	6	PCT-US94-08000-9 Sequence 9, Appl
30	12.4	68.9	28	6	PCT-US92-09202-4 Sequence 4, Appl
31	12.4	68.9	41	6	PCT-US92-09202-3 Sequence 3, Appl
32	12.4	68.9	54	1	US-08-319-836B-1 Sequence 1, Appl
33	12.4	68.9	54	6	PCT-US92-09202-2 Sequence 2, Appl
34	12.4	68.9	54	6	PCT-US95-13142-1 Sequence 1, Appl
35	12.4	68.9	62	1	US-08-319-836B-10 Sequence 10, Appl
36	12.4	68.9	62	6	PCT-US95-13142-10 Sequence 10, Appl
37	12.4	68.9	63	1	US-08-319-836B-11 Sequence 11, Appl
38	12.4	68.9	63	6	PCT-US95-13142-11 Sequence 11, Appl
39	12.4	68.9	85	1	US-08-276-452A-65 Sequence 65, Appl
40	12.4	68.9	85	3	US-08-798-744-65 Sequence 65, Appl
41	12.2	67.8	17	4	US-08-529-878B-35 Sequence 35, Appl
42	12.2	67.8	26	1	US-08-149-096A-3 Sequence 3, Appl
43	12.2	67.8	40	2	US-08-390-878-6 Sequence 6, Appl
44	12.2	67.8	40	2	US-08-390-878-9 Sequence 9, Appl
45	12.2	67.8	40	3	US-08-425-684-64 Sequence 64, Appl

## ALIGNMENTS

RESULT 1  
US-08-123-702-16  
Sequence 16, Application US/08123702  
Patent No. 5604131  
GENERAL INFORMATION:  
APPLICANT: Wadsworth, Samuel  
APPLICANT: Snyder, Benjamin  
APPLICANT: Reddy, Vermuri, B.  
APPLICANT: Wei, Chamer  
TITLE OF INVENTION: A cDNA Genomic Hybrid Sequence Encoding Ap770  
Patent No. 5604131  
TITLE OF INVENTION: Containing a Genomic DNA Insert of the KI and OX-2 Regions  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS: 45  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
City: Atlanta  
STATE: GA  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/123,702  
FILING DATE: 17-SEPT-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: TS1121  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404)-873-8794  
TELEFAX: (404)-873-8795  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 90 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-123-702-16  
Query Match 80.0%; Score 14.4; DB 1; Length 90;  
Best Local Similarity 93.8%; Pred. No. 1.2e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 tgggtgtgtgctt 18  
||||| |||||  
Db 33 TGTGTGTGTGCTT 48

## RESULT 2

US-08-145-704-5  
; Sequence 5, Application US/08145704  
; Patent No. 5567604  
; GENERAL INFORMATION:  
; APPLICANT: Rando, Robert F.  
; APPLICANT: Rannewald, Susan  
; APPLICANT: Zendegeul, Joseph G.  
; APPLICANT: Joshua O. Ojwang  
; TITLE OF INVENTION: Anti-Viral Guanosine-Rich  
; TITLE OF INVENTION: Oligonucleotides  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fulbright & Jaworski  
; STREET: 1301 McKinney, Suite 5100  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: U.S.A.  
; ZIP: 77010-3095  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/145,704  
; FILING DATE: 28-OCT-1993  
; CLASSIFICATION: 514  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/053,027  
; FILING DATE: 23-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Paul, Thomas D.  
; REGISTRATION NUMBER: 32,714  
; REFERENCE/DOCKET NUMBER: D-5574-CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 713/651-5151  
; TELEFAX: 713/651-5246  
; TELEX: 762829  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 36 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-145-704-5

Query Match 76.7%; Score 13.8; DB 1; Length 36;  
Best Local Similarity 88.2%; Pred. No. 2.1e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggtgtgtgtgtgctt 17  
||||| ||||| ||  
Db 9 GGTGTGTGTGTGCTT 25

## RESULT 3

US-08-145-704-7  
; Sequence 7, Application US/08145704  
; Patent No. 5567604  
; GENERAL INFORMATION:  
; APPLICANT: Rando, Robert F.  
; APPLICANT: Rannewald, Susan  
; APPLICANT: Zendegeul, Joseph G.  
; APPLICANT: Joshua O. Ojwang

; TITLE OF INVENTION: Anti-Viral Guanosine-Rich  
; TITLE OF INVENTION: Oligonucleotides  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fulbright & Jaworski  
; STREET: 1301 McKinney, Suite 5100  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: U.S.A.  
; ZIP: 77010-3095  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/145,704  
; FILING DATE: 28-OCT-1993  
; CLASSIFICATION: 514  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/053,027  
; FILING DATE: 23-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Paul, Thomas D.  
; REGISTRATION NUMBER: 32,714  
; REFERENCE/DOCKET NUMBER: D-5574-CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 713/651-5151  
; TELEFAX: 713/651-5246  
; TELEX: 762829  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 36 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 36  
; OTHER INFORMATION: /note="phosphorothioate backbone"  
US-08-145-704-7

Query Match 76.7%; Score 13.8; DB 1; Length 36;  
Best Local Similarity 88.2%; Pred. No. 2.1e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggtgtgtgtgtgctt 17  
||||| ||||| ||  
Db 9 GGTGTGTGTGTGCTT 25

## RESULT 4

PCT-US93-07743-7  
; Sequence 7, Application PC/TUS9307743  
; GENERAL INFORMATION:  
; APPLICANT: Jayaraman, Krishna  
; APPLICANT: Vu, Huynh  
; APPLICANT: Zendegeul, Joseph  
; TITLE OF INVENTION: Cholesteryl-Modified Triple Helix  
; TITLE OF INVENTION: Forming Oligonucleotides and uses thereof  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dr. Adler/ Fulbright & Jaworski  
; STREET: 1301 McKinney St Suite 4100  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77010-3095  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/07743  
CLASSIFICATION:  
FILING DATE: 19930817  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/934,065  
FILING DATE: 21-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Adler Ph.D., Benjamin A.  
REGISTRATION NUMBER: 35,423  
REFERENCE/DOCKET NUMBER: D5523  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-651-5587  
TELEFAX: 713-651-5246  
TELEX: 76-2829  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 36 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 36  
OTHER INFORMATION: /note="a NH2 group is attached to  
OTHER INFORMATION: the 3' end"  
PCT-US93-07743-7

Query Match 76.7%; Score 13.8; DB 6; Length 36;  
Best Local Similarity 88.2%; Pred. No. 2.1e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ggtgtgtgtgtgtgctt 17  
||||| 11 11 11  
DB 9 GGTGTGTGTGTGTGTTT 25

RESULT 5  
PCT-US93-07743-11  
Sequence 11, Application PC/TUS9307743  
GENERAL INFORMATION:  
APPLICANT: Jayaraman, Krishna  
APPLICANT: Vu, Huynh  
APPLICANT: Zendequal, Joseph  
TITLE OF INVENTION: Cholesteryl-Modified Triple Helix  
TITLE OF INVENTION: Forming Oligonucleotides and Uses Thereof  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dr. Adler/ Fulbright & Jaworski  
STREET: 1301 McKinney St Suite 4100  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77010-3095  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/07743  
FILING DATE: 19930817  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/934,065  
FILING DATE: 21-AUG-1992  
ATTORNEY/AGENT INFORMATION:

NAME: Adler Ph.D., Benjamin A.  
REGISTRATION NUMBER: 35,423  
REFERENCE/DOCKET NUMBER: D5523  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-651-5587  
TELEFAX: 713-651-5246  
TELEX: 76-2829  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 36 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1  
OTHER INFORMATION: /note="a cholesterol moiety is  
OTHER INFORMATION: attached to the 3' end"  
PCT-US93-07743-11

Query Match 76.7%; Score 13.8; DB 6; Length 36;  
Best Local Similarity 88.2%; Pred. No. 2.1e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ggtgtgtgtgtgtgctt 17  
||||| 11 11 11  
DB 9 GGTGTGTGTGTGTGTTT 25

RESULT 6  
PCT-US93-07743-16  
Sequence 16, Application PC/TUS9307743  
GENERAL INFORMATION:  
APPLICANT: Jayaraman, Krishna  
APPLICANT: Vu, Huynh  
APPLICANT: Zendequal, Joseph  
TITLE OF INVENTION: Cholesteryl-Modified Triple Helix  
TITLE OF INVENTION: Forming Oligonucleotides and Uses Thereof  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dr. Adler/ Fulbright & Jaworski  
STREET: 1301 McKinney St Suite 4100  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77010-3095  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/07743  
FILING DATE: 19930817  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/934,065  
FILING DATE: 21-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Adler Ph.D., Benjamin A.  
REGISTRATION NUMBER: 35,423  
REFERENCE/DOCKET NUMBER: D5523  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-651-5587  
TELEFAX: 713-651-5246  
TELEX: 76-2829  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 36 base pairs

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1
; OTHER INFORMATION: /note="a propanolamine moiety is
; OTHER INFORMATION: attached to the 3' end"
PCT-US93-07743-16
```

```

Query Match          76.7%; Score 13.8; DB 6; Length 36;
Best Local Similarity 88.2%; Pred. No.2.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 ggtgtgtgtgtgtgtc 17
    |||||  |||  ||
DB 9 GGTGTGTGTGTGTGT 25
```

```

RESULT 7
PCT-US96-11786-5
; Sequence 5, Application PC/TUS9611786
; GENERAL INFORMATION:
; APPLICANT: Rando, Robert F.
; APPLICANT: Fennwald, Susan
; APPLICANT: Zendegui, Joseph G.
; APPLICANT: Ojwang, Joshua O.
; APPLICANT: Hogan, Michael E.
; APPLICANT: Pommer, Eyles
; APPLICANT: Mazunder, Abhijit
; TITLE OF INVENTION: Anti-viral Guanosine-Rich
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Conley, Rose & Tayon, P.C.
; STREET: 600 Travis, Suite 1850
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77002-2912
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/11786
; FILING DATE: 17-JULY-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/535,168; 60/001,505; 60/014,007; 60/013,688;
; APPLICATION NUMBER: 60/015,714; 60/016,271
; FILING DATE: 23-OCT-95; 17-JULY-96; 25-MARCH-96; 19-MARCH-96; 23-
; FILING DATE: APRIL-96; 17-APRIL-96
; ATTORNEY/AGENT INFORMATION:
; NAME: McDaniel, C. Steven
; REGISTRATION NUMBER: 33,962
; REFERENCE/DOCKET NUMBER: 1472-06214
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/238-8010
; TELEFAX: 713/238-8008
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 36 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US96-11786-5
```

```

Query Match          76.7%; Score 13.8; DB 6; Length 36;
Best Local Similarity 88.2%; Pred. No.2.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 ggtgtgtgtgtgtc 17
    |||||  |||  ||
DB 9 GGTGTGTGTGTGTGT 25
```

```

RESULT 8
PCT-US96-11786-7
; Sequence 7, Application PC/TUS9611786
; GENERAL INFORMATION:
; APPLICANT: Rando, Robert F.
; APPLICANT: Fennwald, Susan
; APPLICANT: Zendegui, Joseph G.
; APPLICANT: Ojwang, Joshua O.
; APPLICANT: Hogan, Michael E.
; APPLICANT: Pommer, Eyles
; APPLICANT: Mazunder, Abhijit
; TITLE OF INVENTION: Anti-viral Guanosine-Rich
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Conley, Rose & Tayon, P.C.
; STREET: 600 Travis, Suite 1850
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77002-2912
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/11786
; FILING DATE: 17-JULY-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/535,168; 60/001,505; 60/014,007; 60/013,688;
; APPLICATION NUMBER: 60/015,714; 60/016,271
; FILING DATE: 23-OCT-95; 17-JULY-96; 25-MARCH-96; 19-MARCH-96; 23-
; FILING DATE: APRIL-96; 17-APRIL-96
; ATTORNEY/AGENT INFORMATION:
; NAME: McDaniel, C. Steven
; REGISTRATION NUMBER: 33,962
; REFERENCE/DOCKET NUMBER: 1472-06214
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/238-8010
; TELEFAX: 713/238-8008
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 36 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 36
; OTHER INFORMATION: /note="phosphorothioate
; OTHER INFORMATION: backbone"
PCT-US96-11786-7
```

```

Query Match          76.7%; Score 13.8; DB 6; Length 36;
Best Local Similarity 88.2%; Pred. No.2.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ggtgtgtgtgtgtc 17
```

Db 9 GGCTGTGCTGCTGTTT 25

## RESULT 9

US-08-703-755A-4  
; Sequence 4, Application US/08703755A

; Patent No. 5691145

## GENERAL INFORMATION:

APPLICANT: Pitner, Bruce

APPLICANT: Vonk, Glenn P.

APPLICANT: Nadeau, James G.

TITLE OF INVENTION: DETECTION OF NUCLEIC ACIDS USING

TITLE OF INVENTION: G-QUARTETS

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Richard J. Rodrick, Becton Dickinson and

STREET: 1 Becton Drive

CITY: Franklin Lakes

STATE: NJ

COUNTRY: US

ZIP: 07417

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/703,755A

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Fugit, Donna R.

REGISTRATION NUMBER: 32,135

REFERENCE/DOCKET NUMBER: P-3376

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-703-755A-4

Query Match 73.3%; Score 13.2; DB 2; Length 20;

Best Local Similarity 83.3%; Pred. No. 3.6e+02;

Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggtgtgtgtgtgtgttc 18

Db 1 GGTTTGTGTTTGGTTT 18

## RESULT 10

US-08-145-704-16

; Sequence 16, Application US/08145704

; Patent No. 5567604

## GENERAL INFORMATION:

APPLICANT: Rando, Robert F.

APPLICANT: Fennwald, Susan

APPLICANT: Zendegeul, Joseph G.

APPLICANT: Joshua O. Ojwang

TITLE OF INVENTION: Anti-Viral Guanoxine-Rich

TITLE OF INVENTION: Oligonucleotides

NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fulbright & Jaworski

STREET: 1301 McKinney, Suite 5100

CITY: Houston

STATE: Texas

COUNTRY: U.S.A.

ZIP: 77010-3095

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/145,704

FILING DATE: 28-OCT-1993

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/053,027

FILING DATE: 23-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Paul, Thomas D.

REGISTRATION NUMBER: 32,714

REFERENCE/DOCKET NUMBER: D-5574-CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 713/651-5151

TELEFAX: 713/651-5246

TELEX: 762829

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 24 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-145-704-16

Query Match 73.3%; Score 13.2; DB 1; Length 24;

Best Local Similarity 83.3%; Pred. No. 3.7e+02;

Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggtgtgtgtgtgttc 18

Db 2 GGCTGTGCTGCTGTTT 19

## RESULT 11

PCT-US96-11786-16

; Sequence 16, Application PC/TUS9611786

## GENERAL INFORMATION:

APPLICANT: Rando, Robert F.

APPLICANT: Fennwald, Susan

APPLICANT: Zendegeul, Joseph G.

APPLICANT: Ojwang, Joshua O.

APPLICANT: Hogan, Michael E.

APPLICANT: Pommer, Eyles

TITLE OF INVENTION: Anti-Viral Guanoxine-Rich

TITLE OF INVENTION: Oligonucleotides

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: Conley, Rose & Tayon, P.C.

STREET: 600 Travis, Suite 1850

CITY: Houston

STATE: Texas

COUNTRY: U.S.A.

ZIP: 77002-2912

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/11786

FILING DATE: 17-JULY-1996

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/535,168; 60/001,505; 60/014,007; 60/013,688;

APPLICATION NUMBER: 60/015,714; 60/016,271

FILING DATE: 23-OCT-95; 17-JULY-96; 25-MARCH-96; 19-MARCH-96; 23-



SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/197,770  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 93 10 2484.8  
FILING DATE: 17-FEB-1993  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 45 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-197-770-8

Query Match 73.3%; Score 13.2; DB 1; Length 45;  
Best Local Similarity 83.3%; Pred. No. 3.9e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggtgttggttggttt 18  
||| ||||| ||| |||  
Db 15 ggtgatggtgatgattt 32

RESULT 15  
US-08-495-695B-23  
Sequence 23, Application US/08495695B  
Patent No. 5976814  
GENERAL INFORMATION:  
APPLICANT: Bard, Jonathan A.  
APPLICANT: Walker, Mary  
APPLICANT: Brancheck, Theresa  
APPLICANT: Weinsbank, Richard L.  
TITLE OF INVENTION: DNA ENCODING A HUMAN NEUROPEPTIDE  
TITLE OF INVENTION: Y/PEPTIDE YY/PANCREATIC POLYPEPTIDE  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/495,695B  
FILING DATE: 13-Jan-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 44743-A-PCT-US/JFW/JHB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO

ANTI-SENSE: NO  
US-08-495-695B-23

Query Match 73.3%; Score 13.2; DB 4; Length 46;  
Best Local Similarity 83.3%; Pred. No. 3.9e+02;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggtgttggttggttt 18  
||| ||||| ||| |||  
Db 6 ggtgatggtgatgacctt 23

Search completed: June 4, 2000, 16:09:44  
Job time: 28063 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:05:37 ; Search time 1236.38 Seconds  
(without alignments)  
-15.736 Million cell updates/sec

Title: US-09-164-714-14

Perfect score: 20

Sequence: 1 ccccttaaacatccgcac 20

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 356616

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenBank: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_pl1: \*  
8: gb\_pl2: \*  
9: gb\_pr1: \*  
10: gb\_pr2: \*  
11: gb\_pr3: \*  
12: gb\_ro: \*  
13: gb\_sts: \*  
14: gb\_sy: \*  
15: gb\_un: \*  
16: gb\_vl: \*  
17: em\_fun: \*  
18: em\_hum1: \*  
19: em\_hum2: \*  
20: em\_in: \*  
21: em\_om: \*  
22: em\_or: \*  
23: em\_ov: \*  
24: em\_pat: \*  
25: em\_ph: \*  
26: em\_pl: \*  
27: em\_ro: \*  
28: em\_sy: \*  
29: em\_un: \*  
30: em\_un: \*  
31: em\_vl: \*  
32: gb\_hg1: \*  
33: gb\_hg2: \*  
34: gb\_in1: \*  
35: gb\_in2: \*  
36: em\_ba1: \*  
37: em\_ba2: \*  
38: em\_hum3: \*  
39: em\_hum4: \*  
40: gb\_pr4: \*  
41: gb\_hg3: \*  
42: gb\_hg4: \*  
43: gb\_hg5: \*  
44: gb\_hg6: \*

45: gb\_hg7: \*  
46: em\_hg1: \*  
47: em\_hg2: \*  
48: em\_hg3: \*  
49: em\_hum5: \*  
50: gb\_pl3: \*  
51: gb\_pr5: \*  
52: gb\_hg8: \*  
53: gb\_hg9: \*  
54: gb\_hg10: \*  
55: gb\_hg11: \*  
56: gb\_hg12: \*  
57: gb\_hg13: \*  
58: gb\_hg14: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14.2	71.0	50	5 AR032933	AR032933 Sequence
2	14.2	71.0	50	5 129673	129673 Sequence 54
3	14.2	71.0	50	5 191347	191347 Sequence 54
4	13.8	69.0	69	5 A48319	A48319 Sequence 29
5	13.8	69.0	69	5 A48320	A48320 Sequence 30
6	13.6	68.0	55	1 T1PCT7KBA	K02338 T1 plasmid
7	13.2	66.0	76	10 HS030191	U30191 Human clone
8	12.6	63.0	22	5 AR066834	AR066834 Sequence
9	12.6	63.0	22	5 AR066998	AR066998 Sequence 9
10	12.6	63.0	33	5 A56675	A56675 Sequence 9
11	12.6	63.0	33	5 AR051609	AR051609 Sequence
12	12.2	61.0	46	5 AR032696	AR032696 Sequence
13	12.2	61.0	46	5 129436	129436 Sequence 30
14	12.2	61.0	46	5 191110	191110 Sequence 30
15	12.2	61.0	57	34 AF039609	AF039609 Drosophila
16	12.2	61.0	63	5 A18114	A18114 Oligonucleo
17	12.2	60.0	49	5 156706	156706 Sequence 22
18	12.2	60.0	50	5 AR032926	AR032926 Sequence
19	12.2	60.0	50	5 129666	129666 Sequence 53
20	12.2	60.0	50	5 191340	191340 Sequence 53
21	12.2	60.0	62	10 HCOL4A6S12	U46370 Human type
22	12.2	60.0	72	9 HUMCOL4S11	D63526 Homo sapien
23	12.2	60.0	80	5 AR051980	AR051980 Sequence
24	12.2	60.0	80	5 152154	152154 Sequence 64
25	12.2	60.0	81	16 AF109762	AF109762 Hepatit
26	12.2	60.0	94	13 HSPR33H11	AL009835 H. sapien
27	11.8	59.0	20	5 181222	181222 Sequence 6
28	11.8	59.0	23	5 AR055654	AR055654 Sequence
29	11.8	59.0	89	5 AR011127	AR011127 Sequence
30	11.6	58.0	20	5 AR066722	AR066722 Sequence
31	11.6	58.0	21	5 AR031094	AR031094 Sequence
32	11.6	58.0	24	5 A99109	A99109 Sequence 13
33	11.6	58.0	30	5 A14209	A14209 Oligonucleo
34	11.6	58.0	48	5 AR067776	AR067776 Sequence
35	11.6	58.0	48	5 E13195	E13195 Oligonucleo
36	11.6	58.0	59	5 E13419	E13419 Synthetic o
37	11.6	58.0	59	5 E13434	E13434 Synthetic o
38	11.6	58.0	60	5 AR006756	AR006756 Sequence
39	11.6	58.0	60	5 171268	171268 Sequence 6
40	11.6	58.0	78	5 109608	109608 Sequence 4
41	11.6	58.0	89	16 XCU70893	U70893 Xestia c-ni
42	11.6	58.0	94	11 HUMTL75	M29052 Human inter
43	11.4	57.0	15	5 177853	177853 Sequence 56
44	11.4	57.0	23	24 E09203	E09203 Primer for
45	11.4	57.0	39	12 MUSBMP241	D45010 Mouse gene

#### ALIGNMENTS

RESULT 1  
LOCUS AR032933 50 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 545 from patent US 5869241.  
ACCESSION AR032933  
VERSION AR032933.1 GI:5948538  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE  
1 (bases 1 to 50)  
AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.  
TITLE Method of determining DNA sequence preference of a DNA-binding molecule  
JOURNAL Patent: US 5869241-A 545 09-FEB-1999;  
FEATURES Location/Qualifiers  
source 1..50  
BASE COUNT 13 a 7 c 16 g 14 t  
ORIGIN

Query Match 71.0%; Score 14.2; DB 5; Length 50;  
Best Local Similarity 84.2%; Pred. No. 2.8e+03;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccccttaaacatcgcca 19  
||||| ||| ||| |||  
DB 28 CCCCTTTAAATATCCCA 10

RESULT 2  
LOCUS I29673 50 bp DNA PAT 06-FEB-1997  
DEFINITION Sequence 545 from patent US 5578444.  
ACCESSION I29673  
VERSION I29673.1 GI:1820464  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE  
1 (bases 1 to 50)  
AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M., Turin,L.M. and Fry,K.E.  
TITLE Sequence-directed DNA-binding molecules compositions and methods  
JOURNAL Patent: US 5578444-A 545 26-NOV-1996;  
FEATURES Location/Qualifiers  
source 1..50  
BASE COUNT 13 a 7 c 16 g 14 t  
ORIGIN

Query Match 71.0%; Score 14.2; DB 5; Length 50;  
Best Local Similarity 84.2%; Pred. No. 2.8e+03;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccccttaaacatcgcca 19  
||||| ||| ||| |||  
DB 28 CCCCTTTAAATATCCCA 10

RESULT 3  
LOCUS I91347 50 bp DNA PAT 01-DEC-1998  
DEFINITION Sequence 545 from patent US 5726014.  
ACCESSION I91347  
VERSION I91347.1 GI:3935817  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE  
1 (bases 1 to 50)  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source  
BASE COUNT 19 a 18 c 9 g 23 t

AUTHORS Edwards,C.A., Cantor,C.R., Andrews,B.M. and Turin,L.M.  
TITLE Screening assay for the detection of DNA-binding molecules  
JOURNAL Patent: US 5726014-A 545 10-MAR-1998;  
FEATURES Location/Qualifiers  
source 1..50  
BASE COUNT 13 a 7 c 16 g 14 t  
ORIGIN

Query Match 71.0%; Score 14.2; DB 5; Length 50;  
Best Local Similarity 84.2%; Pred. No. 2.8e+03;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccccttaaacatcgcca 19  
||||| ||| ||| |||  
DB 28 CCCCTTTAAATATCCCA 10

RESULT 4  
LOCUS A48319 69 bp DNA PAT 07-MAR-1997  
DEFINITION Sequence 29 from Patent WO9601900.  
ACCESSION A48319  
VERSION A48319.1 GI:2302112  
KEYWORDS  
SOURCE  
ORGANISM unidentified.  
REFERENCE  
1 (bases 1 to 69)  
AUTHORS Layton,G.T., Garcia-Valcarcel,M., Fowler,W.J. and Harper,D.R.  
TITLE IMMUNODOMINANT POLYPEPTIDES  
JOURNAL Patent: WO 9601900-A 29 25-JAN-1996;  
COMMENT BRITISH BIOTECH PHARM (GB)  
Other publication AU 2803595 960209.  
FEATURES Location/Qualifiers  
source 1..69  
BASE COUNT 23 a 9 c 18 g 19 t  
ORIGIN

Query Match 69.0%; Score 13.8; DB 5; Length 69;  
Best Local Similarity 88.2%; Pred. No. 4.6e+03;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ccccttaaacatcgcc 18  
||||| ||| ||| |||  
DB 64 CCCCTGAAACATCAC 48

RESULT 5  
LOCUS A48320 69 bp DNA PAT 07-MAR-1997  
DEFINITION Sequence 30 from Patent WO9601900.  
ACCESSION A48320  
VERSION A48320.1 GI:2302113  
KEYWORDS  
SOURCE  
ORGANISM unidentified.  
REFERENCE  
1 (bases 1 to 69)  
AUTHORS Layton,G.T., Garcia-Valcarcel,M., Fowler,W.J. and Harper,D.R.  
TITLE IMMUNODOMINANT POLYPEPTIDES  
JOURNAL Patent: WO 9601900-A 30 25-JAN-1996;  
COMMENT BRITISH BIOTECH PHARM (GB)  
Other publication AU 2803595 960209.  
FEATURES Location/Qualifiers  
source 1..69  
BASE COUNT 19 a 18 c 9 g 23 t

ORIGIN

Query Match 69.0%; Score 13.8; DB 5; Length 69;  
Best Local Similarity 88.2%; Pred. No. 4.6e+03;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ccccttaaacatgcgc 18  
||||| ||||| |||  
DB 10 CCCTTGAAACATCACCC 26

RESULT 6  
TIPCT7KBA 55 bp DNA BCT 13-MAR-1996  
LOCUS T1 plasmid (from A.tumefaciens) T-region .7 kb transcript gene, 5'  
DEFINITION flank.  
ACCESSION K02338  
VERSION K02338.1 GI:154732  
KEYWORDS  
SOURCE Plasmid T1 (clone: pT1a6.) DNA.  
ORGANISM plasmid T1  
REFERENCE 1 (bases 1 to 55)  
AUTHORS McPherson,J.C.  
TITLE DNA sequence analysis of crown gall tumor T-DNA encoding the 0.7 kb transcript  
JOURNAL Nucleic Acids Res. 12 (5), 2317-2325 (1984)  
MEDLINE 84169535  
FEATURES  
Source Location/Qualifiers  
1..55  
/organism="Plasmid T1"  
/plasmid="Plasmid T1"  
/specific\_host="Agrobacterium tumefaciens"  
/db\_xref="taxon:2512"  
/clone="pT1a6."

BASE COUNT 14 a 23 c 9 g 9 t

ORIGIN

Query Match 68.0%; Score 13.6; DB 1; Length 55;  
Best Local Similarity 80.0%; Pred. No. 5.9e+03;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ccccttaaacatgcgcac 20  
||||| ||||| |||||  
DB 7 CCCCCGAAATCATCGCCAC 26

RESULT 7  
HSU30191 76 bp mRNA PRI 20-SEP-1996  
LOCUS Human clone 1881-5 T-cell receptor delta chain mRNA, partial cds.  
DEFINITION U30191  
ACCESSION U30191.1 GI:940079  
VERSION  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 76)  
AUTHORS Maeruer,M.J., Martin,D., Walter,W., Liu,K., Zltvogel,L., Lotze,M.T., Haluszcak,K., Rabinowich,H., Duguesnoy,R., Storkus,W. and Human intestinal Vdelta+ lymphocytes recognize tumor cells of epithelial origin  
JOURNAL J. Exp. Med. 183 (4), 1681-1696 (1996)  
MEDLINE 96261669  
REFERENCE 2 (bases 1 to 76)  
AUTHORS Martin,D.M.  
TITLE Direct Submision  
JOURNAL Submitted (23-JUN-1995) Dina M. Martin, Surgery, Biochemistry, & Molecular Genetics, University of Pittsburgh, Desoto and O'Hara

FEATURES Streets, Pittsburgh, PA 15261, USA  
source Location/Qualifiers  
1..76  
/organism="Homo sapiens"  
/isolate="patient #1881"  
/db\_xref="taxon:9606"  
/clone="1881-5"  
/sex="male"  
/cell\_type="cytotoxic T-lymphocyte"  
/tissue\_type="colon tumor"  
/dev\_stage="adult"  
<1..>76  
/codon\_start=2  
/product="T-cell receptor delta chain"  
/protein\_id="AAB08496.1"  
/db\_xref="GI:940080"  
/translation="SWGTGQFLQIKNGVDPVKGQYDKL"

BASE COUNT 20 a 15 c 26 g 15 t

ORIGIN

Query Match 66.0%; Score 13.2; DB 10; Length 76;  
Best Local Similarity 83.3%; Pred. No. 9.8e+03;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccccttaaacatgcgc 18  
||||| ||||| |||||  
DB 58 CCCCTTACAGTATCCGC 41

RESULT 8  
AR066834 22 bp DNA PAT 29-SEP-1999  
LOCUS Sequence 182 from patent US 5851760.  
DEFINITION AR066834  
ACCESSION AR066834  
VERSION AR066834.1 GI:5998056  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 22)  
AUTHORS Evans,G.A. and Smith,M.W.  
TITLE Method for generation of sequence sampled maps of complex genomes  
JOURNAL Patent: US 5851760-A 182 22-DEC-1998;  
FEATURES  
source Location/Qualifiers  
1..22  
/organism="unknown"

BASE COUNT 6 a 2 c 8 g 6 t

ORIGIN

Query Match 63.0%; Score 12.6; DB 5; Length 22;  
Best Local Similarity 78.9%; Pred. No. 2e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ccccttaaacatgcgca 19  
||||| ||||| |||  
DB 22 CCCCTTCAAAACAGTGCTA 4

RESULT 9  
AR066998 22 bp DNA PAT 29-SEP-1999  
LOCUS Sequence 346 from patent US 5851760.  
DEFINITION AR066998  
ACCESSION AR066998.1 GI:5998220  
VERSION  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 22)  
AUTHORS Evans,G.A. and Smith,M.W.  
TITLE Method for generation of sequence sampled maps of complex genomes

JOURNAL Patent: US 5851760-A 346 22-DEC-1998;  
FEATURES Location/Qualifiers  
source 1..22  
BASE COUNT 4 a 4 c 7 g 7 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 5; Length 22;  
Best Local Similarity 78.9%; Pred. No. 2e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ccccttaaacatgccac 20  
1 | | | | | | | | | |  
Db 21 CTCCTTAGAACATAGCCAC 3

RESULT 10  
A56675 33 bp DNA PAT 03-MAR-1998  
LOCUS  
DEFINITION Sequence 9 from Patent EP0733704.  
ACCESSION A56675  
VERSION A56675.1 GI:3712718  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 33)  
AUTHORS Boeck, A.P., Mayer, D., Schlensoy, V.D. and Candussio, A.D.  
TITLE Controllable expression system  
JOURNAL Patent: EP 0733704-A 9 25-SEP-1996;  
CONSORTIUM ELEKTROCHEM IND (DE)  
COMMENT Other publication FI 961323 960925  
Other publication JP 8266291 961015  
Other publication DE 19514056 961017  
Other publication CA 2171668 960925.  
FEATURES Location/Qualifiers  
source 1..33  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 11 a 11 c 7 g 4 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 5; Length 33;  
Best Local Similarity 78.9%; Pred. No. 2e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ccccttaaacatgccca 19  
1 | | | | | | | | | |  
Db 7 CCCGCTTAACATCTCCA 25

RESULT 11  
AR051609 33 bp DNA PAT 29-SEP-1999  
LOCUS  
DEFINITION Sequence 9 from patent US 5830692.  
ACCESSION AR051609  
VERSION AR051609.1 GI:5974973  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 33)  
AUTHORS Bock, A., Mayer, D., Schlensoy, V. and Candussio, A.  
TITLE Expression system which can be regulated  
JOURNAL Patent: US 5830692-A 9 03-NOV-1998;  
FEATURES Location/Qualifiers  
source 1..33  
/organism="unknown"

BASE COUNT 11 a 11 c 7 g 4 t  
ORIGIN

Query Match 63.0%; Score 12.6; DB 5; Length 33;  
Best Local Similarity 78.9%; Pred. No. 2e+04;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ccccttaaacatgccca 19  
1 | | | | | | | | | |  
Db 7 CCCGCTTAACATCTCCA 25

RESULT 12  
AR032696 46 bp DNA PAT 29-SEP-1999  
LOCUS  
DEFINITION Sequence 308 from patent US 5869241.  
ACCESSION AR032696  
VERSION AR032696.1 GI:5948301  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 46)  
AUTHORS Edwards, C.A., Cantor, C.R., Andrews, B.M., Turin, L.M. and Fry, K.E.  
TITLE Method of determining DNA sequence preference of a DNA-binding molecule  
JOURNAL Patent: US 5869241-A 308 09-FEB-1999;  
FEATURES Location/Qualifiers  
source 1..46  
/organism="unknown"

BASE COUNT 8 a 19 c 12 g 7 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 5; Length 46;  
Best Local Similarity 82.4%; Pred. No. 3.4e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccccttaaacatgcc 17  
1 | | | | | | | | | |  
Db 17 CCCCTTAAGCGCGC 33

RESULT 13  
I29436 46 bp DNA PAT 06-FEB-1997  
LOCUS  
DEFINITION Sequence 308 from patent US 5578444.  
ACCESSION I29436  
VERSION I29436.1 GI:1820227  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 46)  
AUTHORS Edwards, C.A., Cantor, C.R., Andrews, B.M., Turin, L.M. and Fry, K.E.  
TITLE Sequence-directed DNA-binding molecules compositions and methods  
JOURNAL Patent: US 5578444-A 308 26-NOV-1996;  
FEATURES Location/Qualifiers  
source 1..46  
/organism="unknown"

BASE COUNT 8 a 19 c 12 g 7 t  
ORIGIN

Query Match 61.0%; Score 12.2; DB 5; Length 46;  
Best Local Similarity 82.4%; Pred. No. 3.4e+04;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccccttaaacatgcc 17  
1 | | | | | | | | | |  
Db 17 CCCCTTAAGCGCGC 33

RESULT 14  
191110

```

LOCUS       191110               46 bp    DNA             PAT       01-DEC-1998
DEFINITION  Sequence 308 from patent US 5726014.
ACCESSION   191110
VERSION     191110.1  GI:3935580
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 46)
AUTHORS    Edwards,C.A., Cantor,C.R., Andrews,B.M. and Turin,L.M.
TITLE      Screening assay for the detection of DNA-binding molecules
JOURNAL    Patent: US 5726014-A 308 10-MAR-1998;
FEATURES
             Location/Qualifiers
             source             1..46
                                 /organism="unknown"
BASE COUNT   8 a          19 c          12 g          7 t
ORIGIN
Query Match          61.0%; Score 12.2; DB 5; Length 46;
Best Local Similarity 82.4%; Pred. No. 3.4e+04;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1  ccccttaaacatgc 17
        |||||
Db      17  CCCCTTTAAAGCGCGC 33

RESULT  15
AF039609  57 bp    DNA             INV       23-MAR-1998
LOCUS     AF039609/c
DEFINITION Drosophila capricorni xanthine dehydrogenase gene, Intron II.
ACCESSION AF039609
VERSION   AF039609.1  GI:2981324
KEYWORDS
SOURCE    Drosophila capricorni.
ORGANISM  Drosophila capricorni.
           Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
           Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
           Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 57)
AUTHORS   Tarrío,R., Rodríguez-Trelles,F. and Ayala,F.J.
TITLE     New drosophila introns originate by duplication
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (4), 1658-1662 (1998)
MEDLINE   98132647
REFERENCE 2 (bases 1 to 57)
AUTHORS   Tarrío,R., Rodríguez-Trelles,F. and Ayala,F.J.
TITLE     Direct Submission
JOURNAL   Submitted (22-DEC-1997) Genetika, Universitat Autònoma de
           Barcelona, Barcelona, CA 08193, Spain
FEATURES
             Location/Qualifiers
             source             1..57
                                 /organism="Drosophila capricorni"
                                 /db_xref="taxon:46840"
                                 1..57
                                 /note="Intron II; xanthine dehydrogenase"
BASE COUNT   16 a          16 c          12 g          21 t
ORIGIN
Intron
Query Match          61.0%; Score 12.2; DB 34; Length 57;
Best Local Similarity 82.4%; Pred. No. 3.4e+04;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      4  cttaaacatgcgcac 20
        | ||||| |||||
Db     21  CATTAAAGATCCGAC 5

```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:24:11 ; Search time 322.35 Seconds  
(without alignments)  
15.523 Million cell updates/sec

Title: US-09-164-714-14

Perfect score: 20  
Sequence: 1 ccccttaaacatcgccac 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 431286

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	14.2	71.0	50	1	Q69795	Epstein Barr Virus
2	14.2	71.0	50	1	T64257	EBV R2 (start site
3	14.2	71.0	50	1	X17545	Test sequence from
4	13.8	69.0	69	1	T11012	Coding sequence from
5	13.6	68.0	42	1	X33064	Human BRCA2 gene P
6	13.4	67.0	36	1	X04254	MMP-7 linker regio
7	13	65.0	21	1	Q41535	Antisense oligomer
8	12.8	64.0	50	1	X15997	PCR primer C1-7 us
9	12.8	64.0	50	1	X16000	PCR primer C2-2 us
10	12.8	64.0	50	1	Q90839	Wild type hMLH1 ge
11	12.6	63.0	22	1	Q82182	Chromosome 11 (loc
12	12.6	63.0	22	1	Q82346	Chromosome 11 (loc
13	12.6	63.0	33	1	T33684	Primer used to amp
14	12.4	62.0	60	1	V09404	Human ICE LAP-6 cd
15	12.2	61.0	28	1	X02116	Human FEN-1 PCR pr
16	12.2	61.0	46	1	Q69558	Human thrombospond
17	12.2	61.0	46	1	T64020	Human thrombospond
18	12.2	61.0	46	1	X17308	Test sequence from
19	12	60.0	49	1	Q99620	Human Ah receptor
20	12	60.0	50	1	Q69788	Epstein Barr Virus
21	12	60.0	50	1	T64250	EBV LI 1 kb gene (
22	12	60.0	50	1	X17538	Test sequence from
23	12	60.0	66	1	V09407	Human ICE LAP-6 cd
24	12	60.0	80	1	Q92534	Forward primer AFI
25	11.8	59.0	20	1	T95216	PCR primer 1 for t
26	11.8	59.0	25	1	V37959	ECCEPO section 1 co
27	11.8	59.0	25	1	V38003	SCCEPO section 1 co
28	11.8	59.0	75	1	V75933	Staphylococcus aur
29	11.6	58.0	20	1	Q82070	Chromosome 11 (loc
30	11.6	58.0	21	1	Q57800	Primer pair 10B D3
31	11.6	58.0	24	1	X27972	PCR primer for B18
32	11.6	58.0	34	1	X27834	Sense primer to am
33	11.6	58.0	48	1	V02491	Bridge oligomer MP
34	11.6	58.0	48	1	V02470	Primer MP92 for a

35	11.6	58.0	51	1	Q87671	Human superoxide d
36	11.6	58.0	51	1	V73847	Human SOD1 derived
37	11.6	58.0	58	1	T94059	Oligo used in cons
38	11.6	58.0	58	1	T94061	Oligo used in cons
39	11.6	58.0	60	1	T95110	Monkey Immunoglobu
40	11.6	58.0	74	1	T63629	Probe for L-prolin
41	11.6	58.0	78	1	003124	DNA sequence of th
42	11.4	57.0	15	1	T57364	RSV N hammerhead r
43	11.4	57.0	23	1	Q74017	Human Interleukin
44	11.4	57.0	25	1	V42672	BCSG1 gene 5' prim
45	11.4	57.0	31	1	X38898	Human genomic DNA

## ALIGNMENTS

RESULT 1	069795/c	069795 standard; DNA; 50 BP.
ID	069795;	
AC	06-MAR-1995 (first entry)	
DE	Epstein Barr virus R2 (start site 88897), target region.	
KW	DNA protein-binding assay; test sequence; screening sequence;	
KM	promoter; target; TATA box; Herpes Simplex Virus; HSV;	
KW	origin of replication; UL9; transcription factor; TFIID; ds.	
OS	Synthetic.	
PN	W09414980-A.	
PD	07-JUL-1994.	
PF	20-DEC-1993; U12388.	
PR	23-DEC-1992; US-996783.	
PR	17-SRP-1993; US-123936.	
PA	(GENE-) GENELABS TECHNOLOGIES INC.	
PI	Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;	
DR	WPI; 94-234711/28.	
PT	Sequence-directed DNA-binding molecules - useful in	
PT	pharmaceuticals and as molecular reagents	
PS	Chaim 28; Page 484; 587pp; English.	
CC	A DNA protein-binding assay is provided, useful for screening	
CC	libraries of synthetic or biological cpds. for their ability	
CC	to bind DNA test sequences. The assay is versatile in that any	
CC	number of test sequences can be tested by placing the test sequence	
CC	adjacent to a defined protein-binding screening sequence. Binding	
CC	of mols. to these test sequences changes the binding characteristics	
CC	of the protein mol. to its cognate binding sequence. When such a mol.	
CC	binds the test sequence, the equilibrium of the DNA:protein complexes	
CC	is disturbed, generating changes in the concentration of free DNA probe.	
CC	One application of this method is to eucaryotic general transcription	
CC	factors (e.g. TFIID), where the target region is typically selected	
CC	from DNA sequences adjacent to the binding site for the eucaryotic	
CC	transcription factor. Numerous exemplary test sequences are given:	
CC	the sequences in Q69251-731 and Q69850 correspond to promoter targets	
CC	(typically, TATA box-contg. sites) for human genes and the sequences in	
CC	Q69732-849 correspond to promoter targets for viral genes. The test	
CC	sequences may also be randomly generated. DNA:protein interaction may	
CC	be used for screening purposes, e.g. the Herpes Simplex Virus (HSV)	
CC	origin of replication and UL9 (see Q69851-52, Q69865 and Q69891).	
SQ	Sequence 50 BP; 13 A; 16 G; 14 T;	
Query Match	71.0%;	Score 14.2; DB 1; Length 50;
Best Local Similarity	84.2%;	Pred. No. 88;
Matches 16; Conservative	0;	Mismatches 3; Indels 0; Gaps 0;
OY	1 ccccttaaacatcgcca 19	
DB		
	28 CCCCTTTTAATATCCCA 10	
RESULT 2	T64257/c	
ID	T64257 standard; DNA; 50 BP.	
AC	T64257;	
DR	17-MAR-1997 (first entry)	

DE EBV R2 (start site 88897) TFIID binding site.  
 KM Duplex DNA; target region; binding characteristic; DNA binding protein;  
 KM TFIID; transcription factor; binding site; inhibition; enhance;  
 KM cancer; inherited genetic disorder; ds.  
 OS Epstein-Barr virus.  
 PN US578444-A.  
 PD 26-NOV-1996.  
 PF 27-JUN-1991; 723618.  
 PR 23-DEC-1992; US-723618.  
 PR 17-SEP-1993; US-996783.  
 PR 20-DEC-1993; US-123936.  
 PA (GENE-) GENELABS TECHNOLOGIES INC.  
 PI Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;  
 DR WPI: 97-020402/02.  
 PT Altering binding characteristics of DNA binding proteins to duplex  
 PT DNA - by attaching specific small cpd. to target region close to the  
 PT protein's binding site, useful in treatment of viral disease, cancer  
 PT etc.  
 PS Claim 6; Column 377; 264pp; English.  
 CC The sequences given in T63713-4312 represent duplex DNA's which act  
 CC as target regions in the method of the invention. The method for  
 CC altering the binding characteristics of a DNA-binding protein to duplex  
 CC DNA comprises contacting the duplex DNA with a small molecule which  
 CC binds sequence-specifically to a target region, where, when the small  
 CC molecule is bound to the target region, it is adjacent to, but not  
 CC overlapping by more than 4 bp, a binding site for a DNA-binding protein.  
 CC The small molecule is added at a concentration effective to alter the  
 CC binding of the DNA binding protein, pref. TFIID, to its binding site on  
 CC the duplex DNA. The binding of the small molecule may inhibit or enhance  
 CC the binding of the DNA-binding protein to its binding site. The  
 CC compounds isolated using this method are potentially useful as  
 CC therapeutic agents for treatment of any disease which involves a  
 CC specific DNA sequence, e.g. cancer, or inherited genetic disorders etc.  
 CC The method is suitable for screening large biological or chemical  
 CC libraries and allows determination of sequence-specific and relative  
 CC affinities of known DNA-binding agents for different DNA sequences.  
 CC The design of these duplex DNA's allows a single DNA:protein interaction  
 CC to be used for screening sequence-specific, or preferential, DNA binding  
 CC proteins that recognise almost any possible sequence (see also T49539-  
 CC 74).  
 CC Sequence 50 BP; 13 A; 7 C; 16 G; 14 T;  
 SQ

Query Match 71.0%; Score 14.2; DB 1; Length 50;  
 Best Local Similarity 84.2%; Pred. No. 88;  
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccccttaaacatgcga 19  
 ||||| ||| ||| |||  
 DB 28 CCCCTTTAAATATCCCA 10

RESULT 3  
 X17545/c  
 AC X17545;  
 DT 06-MAY-1999 (first entry)  
 DE Test sequence: DNA-binding molecule; screening sequence; human;  
 KM nucleic acid amplification; target; viral; ds.  
 OS Epstein-Barr virus.  
 PN US5869241-A.  
 PD 09-FEB-1999.  
 PF 07-JUN-1995; 475228.  
 PR 20-DEC-1993; US-171389.  
 PR 27-JUN-1991; US-723618.  
 PR 23-DEC-1992; US-996783.  
 PR 17-SEP-1993; US-123936.  
 PR 07-JUN-1995; US-475228.  
 PA (GENE-) GENELABS TECHNOLOGIES INC.  
 PI Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;  
 DR WPI: 99-152755/13.

PT Determination of DNA sequence preference of a DNA-binding molecule -  
 PT based on inhibition of binding of protein to oligonucleotide  
 PT sequence attached to test sequence  
 PS Claim 3; Columns 379-380; 270pp; English.  
 CC Sequences X17001 to X17600 represent specifically claimed target test  
 CC sequences that are used in the method of the invention of determining the  
 CC DNA sequence preference of a DNA-binding molecule. The method comprises:  
 CC (1) adding a test molecule and a DNA-binding protein to a mixture of  
 CC duplex DNA test oligonucleotides, each of the test oligonucleotides  
 CC having a test sequence adjacent to a screening sequence, where the  
 CC screening sequence binds to the DNA-binding protein with a binding  
 CC affinity that is independent of the DNA sequence of the test sequence,  
 CC and where the mixture of duplex DNA test oligonucleotides includes  
 CC several test sequences; (ii) incubating the test molecule, the mixture of  
 CC duplex DNA test oligonucleotides and the DNA-binding protein for a time  
 CC sufficient to permit binding of the test molecule to test sequences in  
 CC the duplex DNA; (iii) separating unbound test oligonucleotides from test  
 CC oligonucleotides bound to binding protein; (iv) amplifying the unbound  
 CC test oligonucleotides; (v) repeating steps (ii) to (iv); (vi) isolating  
 CC the amplified test oligonucleotides; and (vii) sequencing the isolated  
 CC test oligonucleotides. Test sequences X17001-X17481 and X17600 correspond  
 CC to promoter targets for human genes and test sequences X17482-X17599  
 CC correspond to promoter targets for viral genes.  
 CC Sequence 50 BP; 13 A; 7 C; 16 G; 14 T;  
 SQ

Query Match 71.0%; Score 14.2; DB 1; Length 50;  
 Best Local Similarity 84.2%; Pred. No. 88;  
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccccttaaacatgcga 19  
 ||||| ||| ||| |||  
 DB 28 CCCCTTTAAATATCCCA 10

RESULT 4  
 T11012/c  
 ID T11012 standard; DNA; 69 BP.  
 AC T11012;  
 DT 09-SEP-1996 (first entry)  
 DE Coding sequence for Varicella zoster gE glycoprotein residues 141-161.  
 KM Glycoprotein; gE; VZV; antibody; gB; immunoreactive; immune response;  
 KM infection; diagnosis; therapy; ds.  
 OS Varicella-zoster virus.  
 PI Key location/Qualifiers  
 FH misc-feature 1..4  
 FT /\*tag= a  
 FT /note= "5' overhang"  
 FT complement (69)  
 FT /\*tag= b  
 FT /note= "5' overhang GATC"  
 PN WO9601900-A1.  
 PD 25-JAN-1996.  
 PF 03-JUL-1995; G01566.  
 PR 07-JUL-1994; GB-013751.  
 PA (BRBI-) BRITISH BIOTECH PHARM LTD.  
 PI Fowler WJ, Garcia-valcarcel Munoz-repiso M, Harper DR;  
 PI Layton GT;  
 DR WPI: 96-097630/10.  
 DR P-PSDB: R88536.  
 PT New isolated Varicella zoster gE polypeptide(s) - used to develop  
 PT products for use in vaccines, passive immunisation and diagnosis  
 PT involving VZV infection  
 PS Example 6; Page 37; 47pp; English.  
 CC This sequence represents the coding sequence for residues 141-161 of the  
 CC Varicella-zoster virus (VZV) gE glycoprotein. The encoded sequence (and  
 CC the sequences represented by R88522-R88535 and R88537-R88549) are used to  
 CC create antibodies against the VZV gE glycoprotein. gE is one of six  
 CC glycoproteins encoded by the VZV genome. From these six proteins, gE and  
 CC gB are the major immunoreactive glycoproteins. These six proteins can be  
 CC used for stimulating an immune response against VZV infection. These  
 CC peptides can also be used for determining the presence of anti-VZV gE  
 CC antibodies in a sample, and in the diagnosis of VZV infection. The



CC antibodies against these sequences can be used for passive immunisation  
CC treatment, and in diagnostic applications. This sequence contains the  
CC major VAV GE immunodominant epitope and allows the development of  
CC products which can produce an enhanced and broader immune response.  
SQ Sequence 69 BP; 23 A; 9 C; 18 G; 19 T;

Query Match 69.0%; Score 13.8; DB 1; Length 69;  
Best Local Similarity 88.2%; Pred. No. 1.5e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 cccttaaacatcgcc 18  
||||| ||||||| ||  
Db 64 CCCTTGAACATCACC 48

RESULT 5  
X33064/c  
ID X33064 standard; DNA: 42 BP.  
AC X33064:  
DT 21-JUN-1999 (first entry)  
DE Human BRCA2 gene PCR primer SEQ ID NO:77.  
KW Human; BRCA2; genetic testing; protein therapy; haplotype; detection;  
KM gene therapy; breast cancer; ovarian cancer; PCR primer; ss.  
OS Synthetic.  
NS Homo sapiens.  
PN WO9009164-A1.  
PD 25-FEB-1999.  
PE 14-AUG-1998; 016905.  
PR 22-MAY-1998; US-084471.  
PR 15-AUG-1997; US-055784.  
PR 07-NOV-1997; US-064926.  
PR 12-NOV-1997; US-065367.  
PR 01-MAY-1998; US-071715.  
PA (ONCO-) ONCOMED INC.  
PI Eszkandari T, Jackson GM, Murphy PD, Olson SJ, Park M,  
PI Rabin MB, Schryer B, White MB, Yoshikawa M;  
DR WPI: 99-190163/16.  
PT New coding sequence haplotypes of the human BRCA2 gene - used to  
PT develop products for determining susceptibility to, detection and  
PT treatment of breast or ovarian cancer  
PS Example 1; Page 34; 226pp; English.  
CC The present invention describes genomic DNA which contains a BRCA2 gene  
CC where the first 12 nucleotides beginning exon 5 are 5'-TCCGTGTTCTC-3'  
CC as in sequence (I) (see X03249), where nucleotides numbers  
CC 5782-5790 are GTTGTCTT as in sequence (IV) (see X30253), and where the  
CC last 20 nucleotides encoding exon 15 are 5'-CTGCGTGTCTCATTAACAG-3' as  
CC in sequence (II) (see X30251) and the first 20 nucleotides beginning  
CC exon 16 are 5'-CTGTATACGTATGGCCTTTC-3' as in sequence (III) (see X30253).  
CC Products and methods from the present invention can be used for  
CC identifying mutations in the BRCA2 gene leading to predisposition or  
CC higher susceptibility to breast or ovarian cancer. They can also be used  
CC for detection and gene therapy for breast and ovarian cancers. They can  
CC be used in methods for monitoring disease progression, for determining  
CC patients suited for gene and protein replacement progression, or for  
CC detecting the presence or quantifying the amount of a tumour growth  
CC inhibitor following such therapy. The BRCA2 protein, polypeptides, their  
CC functional equivalents, antibodies, and PMS may also be useful in the  
CC study of the characteristics of BRCA2 proteins, such as structure and  
CC function of BRCA2 in oncogenesis or subcellular localisation of BRCA2  
CC protein in normal and cancerous cells. X33001 to X33097 represent  
CC PCR primers used in the amplification of the human BRCA2 gene.  
SQ Sequence 42 BP; 14 A; 5 C; 12 G; 11 T;

Query Match 68.0%; Score 13.6; DB 1; Length 42;

Best Local Similarity 80.0%; Pred. No. 1.8e+02;  
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ccccttaaacatcgccac 20  
||| ||||||| ||| ||  
Db 36 CCACTTAAACACACTAC 17

## RESULT 6

ID X04254/c  
ID X04254 standard; DNA: 36 BP.

AC X04254:  
DT 15-APR-1999 (first entry)

DE MWP-7 linker regions of PAP-217.  
KW Ricin-like toxin; cancer; viral infection; parasitic infection;  
KM linker; B chain; A chain; protease; fungal infection; malaria;  
KW leucocyte proliferation; cytomegalovirus; herpes; hepatitis;  
KM rhinovirus; laryngotracheitis; poliomyelitis; varicella zoster;  
KM cystic fibrosis; multiple sclerosis; ds.

OS Unidentified.  
NS Synthetic.  
PN WO9849311-A2.  
PD 05-NOV-1998.  
PE 30-APR-1998; CA0394.  
PR 29-OCT-1997; US-063715.  
PR 30-APR-1997; US-045148.  
PA (DNOV-) DE NOVO ENZYME CORP.

PI Borgford T;  
DR WPI: 99-009431/01.  
PT New nucleic acid encoding ricin-like toxin with an interchain linker  
PT cleaved by protease - is specific for diseased cells, useful for,  
PT e.g. killing selectively cancer or infected cells

PS Disclosure, Fig 4B; 352pp; English.  
CC The present invention describes new purified and isolated nucleic acids  
CC (I) encoding: (1) the A and B chains of a ricin-like toxin (II); and  
CC (11) a heterologous linker, joining the two chains and including a  
CC cleavage recognition site for a disease-specific protease (III). Also  
CC described are: (1) plasmids or baculovirus transfer vectors that contain  
CC (I); and (2) recombinant protein (IV) consisting of the A and B chains  
CC of (II) joined by the specified linker. (IV), produced by expression of  
CC (I) in host cells, are used to inhibit or kill diseased cells that  
CC produce (III), particularly for treating cancers (e.g. leucocyte  
CC proliferation; cancer of ovary, pancreas, breast or prostate; glioma) or  
CC infections caused by fungi, parasites (e.g. malaria) or viruses (e.g.

CC cytomegalovirus (CMV), herpes, hepatitis, rhinovirus, laryngotracheitis,  
CC poliomyelitis or varicella zoster), also cystic fibrosis and multiple  
CC sclerosis. Alternatively, (I) is used to express (IV) in vivo. (IV) is  
CC toxic specifically for (III)-expressing cells and does not depend for  
CC specificity on a cell-binding component. When used to treat virus-  
CC infected cells, transcytosis and cytotoxicity of (IV) are increased by  
CC retrograde translocation from endoplasmic reticulum to cytoplasm (which  
CC some viruses exploit to avoid immune detection), so selectivity and  
CC safety are further improved. (IV) are not toxic until chain A is  
CC released and this occurs only in target cells. The present sequence  
CC represents a nucleotide sequence from the present invention.  
SQ Sequence 36 BP; 6 A; 7 C; 9 G; 14 T;

Query Match 67.0%; Score 13.4; DB 1; Length 36;  
Best Local Similarity 93.3%; Pred. No. 2.2e+02;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 6 ttaaacatcgccac 20  
||||| ||||||| ||  
Db 35 TTAAACTTCGCCAC 21

## RESULT 7

ID Q41535  
ID Q41535 standard; DNA: 21 BP.

AC Q41535:  
DT 10-AUG-1993 (first entry)

DE Antisense oligomer targeting BHRF2 (ATPase/ori binding) AUG site.  
KW Epstein Barr virus; EBV; hybridisation; antisense modulator;  
KW replication; nasopharyngeal carcinoma; Burkitt lymphoma; Sjogren's;  
KW syndrome; infectious mononucleosis; latent; active; infection; ss.

OS Synthetic.  
PN WO9307882-A.  
PD 29-APR-1993.  
PF 23-OCT-1992; U08989.

PR 25-OCT-1991; US-783605.  
 PI (ISIS-) ISIS PHARM INC.  
 PA Anderson KP, Ecker DJ;  
 DR WPI: 93-152174/18.  
 PT Oligo:nucleotide(s) hybridising with RNA of Epstein Barr virus -  
 for treating active, latent and chronic EBV infections and  
 associated diseases e.g. nasopharyngeal carcinoma, Burkitt's  
 lymphoma  
 PS Claim 1: Page 21: 45pp; English.  
 CC The synthetic peptide is an antisense modulator of Epstein Barr  
 virus and pref. contains at least one phosphorothioate linking gp.  
 CC and modifications in the 2' position. These modifications improve  
 CC penetration into regions of the cell that contain nucleic acid and  
 CC also resistance to degradation by nucleases. The oligonucleotide  
 CC targets the AUG site of BBLF2 (Arapase/ori binding) and hybridises thus  
 CC inhibiting replication of EBV. The oligomer may be used for  
 CC treating or preventing EBV-associated diseases, e.g. nasopharyngeal  
 CC carcinoma, Burkitt's lymphoma, Sjogren's syndrome, infectious  
 CC mononucleosis etc. The oligomer is effective against both latent and  
 CC active EBV infection.  
 CC See also Q40575-9 and Q41517-44.  
 SQ Sequence 21 BP; 6 A; 7 C; 3 G; 5 T;

Query Match 65.0%; Score 13; DB 1; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 ccttaaacatc 15  
 |||||  
 DB 1 CCTTAAACATC 13

RESULT 8  
 ID X15997/c  
 AC X15997; standard; DNA; 50 BP.  
 DT 14-MAY-1999 (first entry)  
 DE PCR primer C1-7 used for construction of a synthetic FAOD-P cDNA.  
 KW Fructosylamino acid oxidase; FAOD-P; Penicillium janthinellum S-3413;  
 KW PCR primer: ss.  
 OS Synthetic.  
 OS Penicillium janthinellum.  
 PN J11046769-A.  
 PD 23-FEB-1999.  
 PF 05-AUG-1997; 210609.  
 PR 05-AUG-1997; JP-210609.  
 PA (KYOT-) KYOTO DAIRICH KAGAKU KK.  
 DR WPI: 99-208112/18.  
 PT New DNA coding fructosylamino acid oxidase - synthetically designed  
 PT and recombinantly prepared  
 PS Example 1: Fig 4; 21pp; Japanese.  
 CC PCR primers X15951-6006 were used to construct a synthetic cDNA  
 CC sequence encoding a fructosylamino acid oxidase (FAOD-P). The  
 CC wild type FAOD-P sequence is derived from Penicillium  
 CC janthinellum S-3413.  
 SQ Sequence 50 BP; 15 A; 8 C; 14 G; 13 T;

Query Match 64.0%; Score 12.8; DB 1; Length 50;  
 Best Local Similarity 87.5%; Pred. No. 4.6e+02;  
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 5 tttaaacatgcac 20  
 |||||  
 DB 32 TTTAAAGCATCACAC 17

RESULT 9  
 ID X16000  
 AC X16000; standard; DNA; 50 BP.  
 DT 14-MAY-1999 (first entry)

DE PCR primer C2-2 used for construction of a synthetic FAOD-P cDNA.  
 KW Fructosylamino acid oxidase; FAOD-P; Penicillium janthinellum S-3413;  
 KW PCR primer: ss.  
 OS Synthetic.  
 OS Penicillium janthinellum.  
 PN J11046769-A.  
 PD 23-FEB-1999.  
 PF 05-AUG-1997; 210609.  
 PR 05-AUG-1997; JP-210609.  
 PA (KYOT-) KYOTO DAIRICH KAGAKU KK.  
 DR WPI: 99-208112/18.  
 PT New DNA coding fructosylamino acid oxidase - synthetically designed  
 PT and recombinantly prepared  
 PS Example 1: Fig 4; 21pp; Japanese.  
 CC PCR primers X15951-6006 were used to construct a synthetic cDNA  
 CC sequence encoding a fructosylamino acid oxidase (FAOD-P). The  
 CC wild type FAOD-P sequence is derived from Penicillium  
 CC janthinellum S-3413.  
 SQ Sequence 50 BP; 12 A; 12 C; 9 G; 17 T;

Query Match 64.0%; Score 12.8; DB 1; Length 50;  
 Best Local Similarity 87.5%; Pred. No. 4.6e+02;  
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 5 tttaaacatgcac 20  
 |||||  
 DB 27 TTTAAAGCATCACAC 42

RESULT 10  
 ID Q90839  
 AC Q90839; standard; cDNA; 92 BP.  
 DT 04-MAR-1996 (first entry)  
 DE Wild type hMLH1 gene exon 6.  
 KW hMLH1; wild type; MutL homologue; cancer diagnosis; mismatch repair;  
 KW tumour; susceptibility; mutation detection; exon 6; ss.  
 OS Homo sapiens.  
 PN W09516793-A1.  
 PD 22-JUN-1995.  
 PF 16-DEC-1994; U14746.  
 PR 17-DEC-1993; US-168877.  
 PR 08-MAR-1994; US-209521.  
 PR 09-DEC-1994; US-352902.  
 PA (DAND ) DANA FARBER CANCER INST INC.  
 PA (UYOR-) UNIV OREGON HEALTH SCI.  
 PI Baker SM, Bollaag RJ, Bronner CE, Kolodner RD, Liskay RM;  
 DR WPI: 95-231583/30.  
 PT Determin. of a mutation in a mutL homologue or gene prod. in a tissue  
 PT - used to diagnose cancer susceptibility, and to identify and  
 PT classify a DNA mismatch-repair-defective tumour  
 PS Claim 10: Fig 4A-2; 16pp; English.  
 CC Q90839 is the wild type hMLH1 (a MutL homologue) gene exon 6. A  
 CC mutation in an analogous segment of a hMLH1 or hPMS1 nucleic acid  
 CC isolated from a subject, can be detected by comparing it with the  
 CC above gene fragment. This method can be used to diagnose cancer  
 CC susceptibility, or to identify and classify a DNA mismatch-repair  
 CC defective tumour.  
 SQ Sequence 92 BP; 34 A; 11 C; 24 G; 23 T;

Query Match 64.0%; Score 12.8; DB 1; Length 92;  
 Best Local Similarity 87.5%; Pred. No. 5e+02;  
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 5 tttaaacatgcac 20  
 |||||  
 DB 14 TTTACAACTAGCCAC 29

RESULT 11  
 ID Q82182/c

Query Match	63.0%;	Score 12.6;	DB 1;	Length 22;
Best Local Similarity	78.9%;	Pred. NO.5.3e+02;		
Matches 15;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;

  

QY	1	cccccttaaacatcgcca 19
Db	22	CCCCCTCAAAACAGCGCTA 4

  

RESULT	12
082346/c	
ID	082346 standard; DNA; 22 BP.
AC	082346:
DT	08-SEP-1995 (first entry)
DE	Chromosome 11 (locus D11S1159) STS primer CSR-54a-tz.
KW	sequence sampled mapping; genomic analysis; complex genome mapping;
RM	cosmid library; chromosome 11; sequence tagged site; STS analysis; ss.
SN	Synthetic.
UN	MO9429486-A.
PD	22-DEC-1994.
PE	15-JUN-1994; U06810.
PR	15-JUN-1993; US-078471.
PR	07-SEP-1993; US-117952.
PA	(SALK ) SALK INST BIOLOGICAL STUDIES.
PI	Evans GA, Smith MW;
DR	WPI: 95-036508/05.

  

PT	Sequencing complex genomes, present as fragments in a cosmid
PT	library - by sequencing end-specific nucleotides of each clone
PT	then correlating with spatial relationship of cosmid, esp. for
PT	mammalian chromosomes.
PS	Example 4, Page 77; 128PP; English.
CC	Sequences were determined from the ends of chromosome 11-specific
CC	cosmids by automated sequencing without intermediate subcloning.
CC	A sample of 371 DNA sequence fragments were determined and of
CC	these, 277 were suitable for STS primer prediction by computer
CC	analysis (using the "Primer" program available from E.lander, MIT).

DT 18-MAY-1998 (first entry)

DE Human ICE LAP-6 cDNA amplifying 3' primer 1.  
KM ICE LAP-6; interleukin-1 beta converting enzyme apoptosis protease-6;  
KW treatment; viral infection; tumour; inflammation; osteoporosis;  
KM AIDS; Alzheimer's disease; human; PCR primer; ss.  
OS Synthetic.  
OS Homo sapiens.  
PN EP-808904-A2.  
PD 26-NOV-1997.  
PF 19-MAY-1997; 303397.  
PR 05-JUN-1996; US-018961.  
PR 20-MAY-1996; US-017949.  
PR 23-MAY-1996; US-020344.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PA (SMK ) SMITHKLINE BEECHAM CORP.  
PI (UNMI ) UNIV MICHIGAN.  
PI Dixit VM, He W, Kikly KK, Ruben SM;  
DR WPI: 98-001790/01.  
PT DNA encoding interleukin-1 beta converting enzyme apoptosis  
PT protease-6 - useful to develop products to treat, e.g. viral  
PT infection, tumour, Alzheimer's disease, inflammation, osteoporosis  
PT and AIDS  
PT Example 1; Page 24; 44pp; English.  
CC This primer is used for the PCR amplification of a human interleukin-1  
CC beta converting enzyme apoptosis protease-6 (ICE LAP-6) cDNA. The ICE  
CC LAP-6 polypeptide and agonists to the polypeptide can be used to induce  
CC apoptosis, e.g. as an antiviral or antitumour agent, control embryonic  
CC development and tissue homeostasis and the roles of such factors in  
CC dysfunction and disease. Antagonists which inhibit the activity of the  
CC ICE LAP-6 polypeptide can be used to treat, Alzheimer's or Parkinson's  
CC disease, rheumatoid arthritis, septic shock, sepsis, stroke, chronic,  
CC acute or central nervous system inflammation, osteoporosis, ischemia  
CC reperfusion injury, cell death associated with cardiovascular disease,  
CC polycystic kidney disease, degenerative liver disease, multiple sclerosis,  
CC cardiovascular disease, degenerative liver disease, multiple sclerosis,  
CC cerebellar degeneration, ischaemic injury, myocardial infarction,  
CC acquired immunodeficiency syndrome (AIDS), myelodysplastic syndrome,  
CC aplastic anaemia, male pattern baldness and head injury damage. They can  
CC also be used for detection and diagnosis.  
SQ Sequence 60 BP; 12 A; 4 C; 21 G; 23 T;

Query Match 62.0%; Score 12.4; DB 1; Length 60;  
Best Local Similarity 92.9%; Pred. No. 7.6e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 ccttaaacatgc 17  
|||||  
DB 43 CTTTAAACATCAC 30

RESULT 15  
X02116  
ID X02116 standard; DNA: 28 BP.  
AC X02116.  
DT 23-APR-1999 (first entry)  
DE Human FEN-1 PCR primer #2.  
KW FEN-1; human; flap endonuclease; detection; diagnosis; carcinogen;  
KW neoplasma; antineoplastic agent; cleavage; PCR primer; ss.  
OS Synthetic.  
OS Homo sapiens.  
PN US874283-A.  
PD 23-FEB-1999.  
PF 30-MAY-1995; 455968.  
PR 30-MAY-1995; US-455968.  
PA (HARR/) HARRINGTON J J.  
PA (HSE/) HSEH C.  
PA (LIEB/) LIEBER M R.  
PI Harrington JJ, Hsieh C, Lieber MR;  
DR WPI: 99-179985/15.  
PT DNA encoding flap endonuclease polypeptides - useful for producing  
PT e.g. recombinant polypeptides  
PS Disclosure: Column 25; 58pp; English.  
CC X02115-X02116 are PCR primers used in the isolation and amplification of

CC a human FEN-1 (flap endonuclease) protein. This protein can be used in  
CC methods for detecting a pathological condition in a patient, for  
CC diagnostic purposes, for screening for antineoplastic agents and  
CC carcinogens, for diagnostic staging of neoplasia, for producing  
CC recombinant flap endonuclease for use as research or diagnostic reagents,  
CC for producing antibodies reactive with the novel polypeptides, for  
CC producing transgenic nonhuman animals expressing the novel polypeptides  
CC encoded by a transgene. The invention also provides novel molecular  
CC cloning techniques and reagents involving cleavage of a flap or nick with  
CC a flap endonuclease.  
SQ Sequence 28 BP; 4 A; 9 C; 1 G; 14 T;

Query Match 61.0%; Score 12.2; DB 1; Length 28;  
Best Local Similarity 82.4%; Pred. No. 8.8e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ccccttaaacatgc 17  
|||||  
DB 9 CCCCTTTTAAACTTCCC 25

Search completed: June 4, 2000, 16:24:14  
Job time: 28910 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 13:53:19 ; Search time 4521.53 Seconds  
(without alignments)  
17.929 Million cell updates/sec

Title: US-09-164-714-14

Perfect score: 20

Sequence: 1 ccccttaaacatcgccac 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4857316 segs, 2026611650 residues

Total number of hits satisfying chosen parameters: 119968

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: gb\_est1: \*  
21: gb\_est2: \*  
22: gb\_est3: \*  
23: gb\_est4: \*  
24: gb\_est5: \*  
25: gb\_est6: \*  
26: gb\_est7: \*  
27: gb\_est8: \*  
28: gb\_est9: \*  
29: gb\_est10: \*  
30: gb\_est11: \*  
31: gb\_est12: \*  
32: gb\_est13: \*  
33: gb\_est14: \*  
34: gb\_est15: \*  
35: gb\_est16: \*  
36: gb\_est17: \*  
37: gb\_est18: \*  
38: gb\_est19: \*  
39: gb\_est20: \*  
40: gb\_est21: \*  
41: gb\_est22: \*  
42: gb\_est23: \*  
43: gb\_est24: \*  
44: gb\_est25: \*

45: gb\_est26: \*  
46: gb\_est27: \*  
47: gb\_est28: \*  
48: gb\_est29: \*  
49: gb\_est30: \*  
50: gb\_est31: \*  
51: gb\_est32: \*  
52: em\_est20: \*  
53: em\_est21: \*  
54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*  
59: gb\_est33: \*  
60: gb\_est34: \*  
61: gb\_est35: \*  
62: gb\_est36: \*  
63: gb\_est37: \*  
64: gb\_est38: \*  
65: em\_est27: \*  
66: em\_est28: \*  
67: em\_est29: \*  
68: em\_est30: \*  
69: gb\_est39: \*  
70: gb\_est40: \*  
71: gb\_est41: \*  
72: gb\_est42: \*  
73: gb\_est43: \*  
74: gb\_est44: \*  
75: em\_est31: \*  
76: em\_est32: \*  
77: em\_est33: \*  
78: em\_est34: \*  
79: gb\_est45: \*  
80: gb\_est46: \*  
81: gb\_est47: \*  
82: gb\_est48: \*  
83: gb\_est49: \*  
84: gb\_est50: \*  
85: gb\_est51: \*  
86: em\_est34: \*  
87: em\_est35: \*  
88: em\_est36: \*  
89: em\_est37: \*  
90: gb\_est52: \*  
91: gb\_est53: \*  
92: gb\_est54: \*  
93: gb\_est55: \*  
94: gb\_est56: \*  
95: em\_est38: \*  
96: em\_est39: \*  
97: em\_est40: \*  
98: em\_est41: \*  
99: em\_est42: \*  
100: em\_est43: \*  
101: em\_est44: \*  
102: gb\_est57: \*  
103: gb\_est58: \*  
104: em\_est45: \*  
105: gb\_est59: \*  
106: gb\_est60: \*  
107: gb\_est61: \*  
108: gb\_est62: \*  
109: gb\_est63: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result

Query

SUMMARIES

No.	Score	Match Length	DB	ID	Description
C 1	13.8	69.0	70	63	A1958884 f405d07.y
C 2	13.4	67.0	49	40	AA985666 o71d08.s
C 3	13.4	67.0	91	46	A1423836 tg35b09.x
C 4	13.2	66.0	57	21	R03760 pk10h10..s1
C 5	13.2	66.0	58	40	AA935194 ny39b06.s
C 6	13.2	66.0	59	84	B00362
C 7	13	65.0	76	37	AA730470 nw47e12.s
C 8	12.8	64.0	70	48	A1568312 tne8e07.x
C 9	12.8	64.0	76	39	AA844635 aj15vd07.s
C 10	12.8	64.0	81	45	A1347508 tC06g07.x
C 11	12.8	64.0	90	36	AA649668 ns55e07.s
C 12	12.8	64.0	92	23	H55177 CHR220116 C
C 13	12.6	63.0	55	47	A1480496 vq34a03.x
C 14	12.6	63.0	84	41	AU014474 AD014474
C 15	12.6	63.0	84	41	AU014499 AU014499
C 16	12.6	63.0	88	36	AA607792 v049e08.r
C 17	12.6	63.0	96	63	A1985538 ws06g08.x
C 18	12.4	62.0	58	46	A1444147 fB44c01.y
C 19	12.4	62.0	83	40	AA961514 oq79a09.s
C 20	12.4	62.0	88	28	AA080361 m687f05.r
C 21	12.2	61.0	85	25	NA63504 yw62c04..s1
C 22	12.2	61.0	85	79	AA248908 2819210.3
C 23	12.2	61.0	87	40	AA916294 on22b10.s
C 24	12.2	61.0	87	81	AA423938 sh58b03.y
C 25	12.2	61.0	98	30	AA210795 zr90f02.r
C 26	12.2	61.0	98	23	H16468 ym22d07..s1
C 27	12.2	61.0	97	52	A1089540 A74bIdops
C 28	12.2	61.0	97	59	A1766526 wv65a06.r
C 29	12.2	61.0	100	30	AA219970 mve5a06.r
C 30	12.2	61.0	100	37	AA679974 ag51d11.s
C 31	12.2	61.0	100	47	A1496780 fB61c03.y
C 32	12	60.0	25	42	A1150986 qb52a01.x
C 33	12	60.0	37	60	A1811359 tw38h04.x
C 34	12	60.0	46	42	A1811359 tw38h04.x
C 35	12	60.0	46	46	A1094098 qe33h02.s
C 36	12	60.0	52	44	A1443574 sa33g10.x
C 37	12	60.0	73	43	A1251834 qu77e05.x
C 38	12	60.0	77	28	A1217742 qn70B04.x
C 39	12	60.0	82	43	AA076864 7B07E01 C
C 40	12	60.0	88	31	A1182437 uc25h09.r
C 41	12	60.0	96	39	AA281273 z594a08.r
C 42	11.8	59.0	71	79	AA822235 vW08e07.r
C 43	11.8	59.0	79	61	AAW278253 sf41f11.y
C 44	11.8	59.0	80	23	A1828343 wR83g04.x
C 45	11.8	59.0	85	36	H55403 CHR220342 C
					AA621082 at34a07.s

RESULT	1	ALIGNMENTS
LOCUS	A1958884/c	
DEFINITION	A1958884 70 bp mRNA EST 20-AUG-1999	
ACCESSION	AF054d07.y1 zebrafish washu	
VERSION	SM:MSH2.HUMAN P43246 DNA MISMATCH REPAIR PROTEIN MSH2. ; mRNA	
KEYWORDS	sequence.	
SOURCE	A1958884	
ORGANISM	A1958884.1 GI:5751597	
REFERENCE	EST.	
AUTHORS	zebrafish.	
	Danio rerio.	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;	
	Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;	
	Cyprinoidae; Cyprinidae; Rasbora; Danio.	
	1 (bases 1 to 70)	
	Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M.,	
	Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,	
	Underwood,K., Steptoe,M., Theisling,B., Allen,M., Bowers,Y.,	
	Parson,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schutk,R.,	
	Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,	
	Waterston,R. and Wilson,R.	

[illegible]



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 49)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)

JOURNAL On Jan 14, 1998 this sequence version replaced gi:1798731.  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)

Trace considered overall poor quality  
Insert Length: 517 Std Error: 0.00  
Seq primer: -40m3 fwd. ET from Amersham  
High quality sequence stop: 1.

## FEATURES

source

1. 49  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1601295"  
/clone.lib="NCI.CGAP.Lu5"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: p773D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "  
BASE COUNT 12 a 7 c 15 g 15 t  
ORIGIN

Query Match 67.0%; Score 13.4; DB 40; Length 49;  
Best Local Similarity 93.3%; Pred. No. 9.1e+03;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 ccccttaaacatcg 16  
|||||  
Db 36 CCCTTATACATCG 22

RESULT 3  
AI423836 91 bp mRNA EST 09-MAR-1999  
LOCUS t93b09.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
DEFINITION IMAGE:2110745 3', mRNA sequence.  
ACCESSION AI423836  
VERSION AI423836.1 GI:4269767  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 91)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Apr 21, 1998 this sequence version replaced gi:3071506.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov  
This clone is available royalty-free through LLNL; contact the Image Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -40UP from Gibco  
High quality sequence stop: 68.

## FEATURES

source

1. 91  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2110745"  
/clone.lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHT, and B-cell NCI CGAP GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo. "  
BASE COUNT 28 a 15 c 18 g 30 t  
ORIGIN

Query Match 67.0%; Score 13.4; DB 46; Length 91;  
Best Local Similarity 93.3%; Pred. No. 8.4e+03;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 6 ttaaacatcgccac 20  
|||||  
Db 58 TTTAAACATCGCCAC 44

RESULT 4  
R03760 57 bp mRNA EST 31-MAR-1995  
LOCUS PK10h10.s1 Kuwabara Mixed stage C. briggsae Caenorhabditis briggsae  
DEFINITION CDNA. mRNA sequence.  
ACCESSION R03760  
VERSION R03760.1 GI:753496  
KEYWORDS EST.  
SOURCE Caenorhabditis briggsae.  
ORGANISM Caenorhabditis briggsae

Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida; Rhabditina; Rhabditioidea; Rhabditidae; Pelodermidae; Caenorhabditis.  
1 (bases 1 to 57)  
Hiller, L., Chiapelli, B., Chissee, S., Clark, N., Couch, J., Dubuque, T., Hawkins, M., Holman, M., Hultman, M., Kuwaba, T., Kuwabara, P., Le, M., Mardis, E., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevasakis, E., Waterston, R., Wohlmann, P. and Wilson, R.  
Washington University Caenorhabditis briggsae EST project  
Unpublished (1995)  
Other ESTs: PK10h10.r1  
Contact: Marra MA

Washington University Genome Sequencing Center  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1455  
Fax: 314 286 1810  
Email: [mmarra@watson.wustl.edu](mailto:mmarra@watson.wustl.edu)  
PCR\_F: TGTAAACGACGCCAGTGACCACTTACGCTCG  
PCR\_B: CAGGAACGCTATGACCTTATGAGATTTCTTCAGGGA  
Source: Washington University Genome Sequencing Center  
PCR amplified DNA is available from Washington University Genome Sequencing Center. Aliquots of the library may be requested from P. Kuwabara ([pekemrc-lmb.cam.ac.uk](mailto:pekemrc-lmb.cam.ac.uk)).  
Insert Length: 1138 Std Error: 0.00  
Seq primer: Commercially available -21M13 dye primer

High quality sequence stop: 249.  
Location/Qualifiers

# FEATURES

source

1..57

/organism="Caenorhabditis briggsae"  
/strain="G16 Gujarat"  
/db\_xref="taxon:6238"

/clone\_lib="Kwabara Mixed stage C. briggsae"  
/note="Vector: Lambda g10; Site\_1: EcoRI; Site\_2: EcoRI;  
Stage:mixed, Sex:hermaphrodite. Library construction:  
First strand oligo(dT) primed. Second strand was as per  
Gubler/Hoffman. Ligated to EcoRI adaptors. Library is  
non-directional. Library is non-normalized. Library  
constructed by P.E. Kwabara. Additional details on  
construction of the library are described in P.E.  
Kwabara and S. Shah, NAR 22: 4414 - 4418 (1994). Adaptor  
sequence: GAATTC CGTTCGTCTCG"

BASE COUNT 12 a 15 c 14 g 15 t 1 others

ORIGIN

## Query Match

Best Local Similarity 66.0%; Score 13.2; DB 21; Length 57;  
Pred. No. 1.1e+04;

Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ccccttaaacatgcgca 19

Db 5 CACTTAAATCATTGCGCA 22

## RESULT 5

LOCUS

AA935194 58 bp mRNA EST 28-APR-1998  
ny99b06.s1 NCI\_CGAP\_Pri2 Homo sapiens cDNA clone IMAGE:1274099  
similar to SW:RL29\_HUMAN P47914 60S RIBOSOMAL PROTEIN L29 ; mRNA  
sequence.

DEFINITION

AA935194.1 GI:3092351

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 58)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

Unpublished (1997)

On Jan 14, 1998 this sequence version replaced gi:1797375.  
Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: W. Douglas Fligg, Ph.D., Paul H. Duray, M.D.,  
Rodrigo F. Chuang, M.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: David B. Krizman, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILM at:

www.bio.linn.gov/dbcp/image/image.html

Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 1.

Location/Qualifiers

1..58

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="1274099"  
/clone\_lib="NCI\_CGAP\_Pri2"  
/sex="male"  
/tissue\_type="metastatic prostate bone lesion"  
/lab\_host="DH10B"  
/note="Vector: PAMPI0. mRNA made from metastatic prostate  
lesion of the bone, cDNA made by oligo-dT priming."

## FEATURES

source

Non-directionally cloned. Size-selected on agarose gel,  
average insert size 600 bp. Library made by D. Krizman,  
NIH."

BASE COUNT 15 a 15 c 15 g 13 t

ORIGIN

Query Match 66.0%; Score 13.2; DB 40; Length 58;  
Best Local Similarity 83.3%; Pred. No. 1.1e+04;

Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ccccttaaacatgcgca 20

Db 13 CCTTAAATCATTGCGAC 30

## RESULT 6

LOCUS

B00362 59 bp DNA GSS 13-JUL-1996  
CSRL-108d12-u CSRL flow sorted Chromosome 11 specific cosmid Homo  
sapiens genomic clone CSRL-108d12, genomic survey sequence.

DEFINITION

B00362

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 59)  
Evans,G.A., Burbee,D., Davies,C., Hahner,L., Oliver,T., Gilbert,M.,  
Jones,D., Ward,T., Gillilan,E., Schagemann,J., Probst,S.,  
Harris,J., DeFord,J., McFarland,J., Burzinski,K., Khan,M.,  
Kupfer,K. and Garner,H.R.  
Genomic Sequence Sampled Map of Chromosome 11  
Unpublished (1996)

TITLE

JOURNAL

COMMENT

Contact: Evans GA, Shane Probst  
Mcdermott Center for Human Growth and Development  
University of Texas Southwestern Medical Center At Dallas  
5323 Harry Hines Blvd, Dallas TX 75235-8591  
Tel: 214-648-1600  
Fax: 214-648-1666  
Email: gervans@utsu.swmed.edu, shane@mcdermott.swmed.edu

Seq primer: 17

Class: cosmid ends

High quality sequence stop: 59.

Location/Qualifiers

1..59

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="CSRL-108d12"  
/clone\_lib="CSRL flow sorted Chromosome 11 specific  
cosmid"  
/sex="female"  
/cell\_type="chimeric hamster somatic cell hybrid"  
/note="Vector: SCOS-1; Human Chromosome 11 specific cosmid  
library prepared from flow sorted human Chromosome 11  
derived from Chinese Hamster Ovary (CHO) monochromosomal  
somatic cell hybrid, J1"

BASE COUNT 17 a 15 c 13 g 13 t 1 others

ORIGIN

## FEATURES

source

1..59

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="CSRL-108d12"  
/clone\_lib="CSRL flow sorted Chromosome 11 specific  
cosmid"  
/sex="female"  
/cell\_type="chimeric hamster somatic cell hybrid"  
/note="Vector: SCOS-1; Human Chromosome 11 specific cosmid  
library prepared from flow sorted human Chromosome 11  
derived from Chinese Hamster Ovary (CHO) monochromosomal  
somatic cell hybrid, J1"

BASE COUNT 17 a 15 c 13 g 13 t 1 others

ORIGIN

Query Match 66.0%; Score 13.2; DB 84; Length 59;  
Best Local Similarity 78.9%; Pred. No. 1.1e+04;

Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ccccttaaacatgcgca 19

Db 34 CCCCTTTAAATATANGCA 16

## RESULT 7

LOCUS

AA730470/c

DEFINITION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 59)  
Evans,G.A., Burbee,D., Davies,C., Hahner,L., Oliver,T., Gilbert,M.,  
Jones,D., Ward,T., Gillilan,E., Schagemann,J., Probst,S.,  
Harris,J., DeFord,J., McFarland,J., Burzinski,K., Khan,M.,  
Kupfer,K. and Garner,H.R.  
Genomic Sequence Sampled Map of Chromosome 11  
Unpublished (1996)

TITLE

JOURNAL

COMMENT

Contact: Evans GA, Shane Probst  
Mcdermott Center for Human Growth and Development  
University of Texas Southwestern Medical Center At Dallas  
5323 Harry Hines Blvd, Dallas TX 75235-8591  
Tel: 214-648-1600  
Fax: 214-648-1666  
Email: gervans@utsu.swmed.edu, shane@mcdermott.swmed.edu

LOCUS AA730470 76 bp mRNA EST 06-JAN-1998  
DEFINITION nm47612.s1 NCI\_CGAP\_Ew1 Homo sapiens CDNA clone IMAGE:1249774  
similar to TR:Q16236 Q16236 NRP2. ; mRNA sequence.  
ACCESSION AA730470  
VERSION AA730470.1 GI:2751674  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiinae; Homo.  
REFERENCE 1 (bases 1 to 76)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced g1:1393484.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,  
Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bdrp/image/image.html](http://www-bio.llnl.gov/bdrp/image/image.html)

FEATURES  
source  
1..76  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1249774"  
/clone\_1lb="NCI\_CGAP\_Ew1"  
/tissue\_type="Ewing's sarcoma"  
/lab\_host="DH10B"  
/note="Vector: PAMP10; mRNA made from Ewing's sarcoma,  
CDNA made by oligo-dT priming. Non-directionally cloned.  
Size-selected on agarose gel. average insert size 600 bp.  
Reference: Krizman et al. (1996) Cancer Research  
56:5380-5383."

BASE COUNT 27 a 16 c 15 g 18 t  
ORIGIN

Query Match 65.0%; Score 13; DB 37; Length 76;  
Best Local Similarity 100.0%; Pred. No. 1.4e+04;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 aaacatcgccac 20  
|||||  
Db 36 AAAACATCGCCAC 24

RESULT 8  
AI568312 70 bp mRNA EST 29-MAR-1999  
LOCUS AI568312/c  
DEFINITION tm68607.x1 NCI\_CGAP\_Lym12 Homo sapiens CDNA clone IMAGE:2174724 3'  
similar to TR:Q01944 Q01944 EXTENSIN; contains MER22.b1 TARI  
repetitive element ; mRNA sequence.  
ACCESSION AI568312  
VERSION AI568312  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiinae; Homo.  
REFERENCE 1 (bases 1 to 70)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL Unpublished (1997)  
COMMENT On Jun 22, 1998 this sequence version replaced g1:3247277.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Life Technologies catalog #: 11547-015  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bdrp/image/image.html](http://www-bio.llnl.gov/bdrp/image/image.html)

FEATURES  
source  
1..70  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2174724"  
/clone\_1lb="NCI\_CGAP\_Lym12"  
/tissue\_type="lymphoma, follicular mixed small and large  
cell"  
/lab\_host="DH10B"  
/note="Organ: lymph node; Vector: pCMV-SPORT6; Site\_1:  
SalI; Site\_2: NotI; Cloned unidirectionally. Primer:  
Oligo dT. Average insert size 1.25 kb. Life Technologies  
catalog #: 11547-015"

BASE COUNT 19 a 13 c 32 g 6 t  
ORIGIN

Query Match 64.0%; Score 12.8; DB 48; Length 70;  
Best Local Similarity 87.5%; Pred. No. 1.8e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ccccttaaacatcg 16  
|||||  
Db 41 CCCCTTAAAAACCG 26

RESULT 9  
AA844655/c 76 bp mRNA EST 31-DEC-1998  
LOCUS AA844655/c  
DEFINITION aj574d03.s1 Soares,Testis\_NHT Homo sapiens CDNA clone IMAGE:1394405  
3', mRNA sequence.  
ACCESSION AA844655  
VERSION AA844655  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiinae; Homo.  
REFERENCE 1 (bases 1 to 76)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 17, 1998 this sequence version replaced g1:2043793.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldio, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bdrp/image/image.html](http://www-bio.llnl.gov/bdrp/image/image.html)

Insert Length: 1467 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham.  
Location/Qualifiers

```
source
1. 76
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1394405"
/clone_1ib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5']
TGTTCACATCTGAGTGGAGCGCCGCCCAATTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
```

BASE COUNT 18 a 17 c 22 g 19 t

ORIGIN

Query Match 64.0%; Score 12.8; DB 39; Length 76;  
Best Local Similarity 87.5%; Pred. No. 1.7e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccccttaaacatcg 16  
||||| |||||

Db 73 CCCCTTGAACCATCG 58

RESULT 10  
A1347508 81 bp mRNA EST 08-APR-1999  
LOCUS A1347508.1  
DEFINITION t06g07.x1 NCI\_CGAP\_Col6 Homo sapiens cDNA clone IMAGE:2063100 3',  
mRNA sequence.  
ACCESSION A1347508  
VERSION A1347508.1 GI:4084714  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 81)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/bbrp/image/image.html

JOURNAL

COMMENT

FEATURES

source

Seq primer: -40bp from Gibco.  
Location/Qualifiers

1. 81  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2063100"  
/clone\_1ib="NCI\_CGAP\_Col6"  
/tissue\_type="Colon tumor, RER+"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI-CGAP\_Col6 was  
prepared, and ss circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(clonoids 1057416-106125, and 114584-1145351).  
Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 18 a 21 c 22 g 20 t

ORIGIN

Query Match 64.0%; Score 12.8; DB 45; Length 81;  
Best Local Similarity 87.5%; Pred. No. 1.7e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccccttaaacatcg 16  
||||| |||||

Db 76 CCCCTTGAACCATCG 61

RESULT 11  
AA649968 90 bp mRNA EST 13-NOV-1997  
LOCUS AA649968.C  
DEFINITION ns55e07.s1 NCI\_CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1187556 3',  
mRNA sequence.  
ACCESSION AA649968  
VERSION AA649968.1 GI:2577296  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 90)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Nov 29, 1993 this sequence version replaced gi:430517.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/bbrp/image/image.html

JOURNAL

COMMENT

FEATURES

source

Insert Length: 191 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham.  
Location/Qualifiers

1. 90  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1187556"  
/clone\_1ib="NCI\_CGAP\_Pr22"  
/sex="male"  
/tissue\_type="normal prostate"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)  
with a modified polylinker; 1st strand cDNA was prepared  
from normal prostate bulk tissue, and was then primed with  
a Not I - oligo(dT) primer. Double-stranded cDNA was  
ligated to Eco RI adaptors (Pharmacia), digested with Not  
I and cloned into the Not I and Eco RI sites of the  
modified pT7T3 vector. Library is normalized, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 23 a 22 c 24 g 21 t

ORIGIN

Query Match 64.0%; Score 12.8; DB 36; Length 90;  
Best Local Similarity 87.5%; Pred. No. 1.7e+04;

Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccccttaaacatcg 16  
||||| |||||  
Db 75 CCCCTTGAACCATCG 60

RESULT 12  
H55177/c 92 bp DNA EST 07-DEC-1995  
LOCUS CHR220116 Chromosome 22 exon Homo sapiens genomic clone C22\_151.5',  
DEFINITION mRNA sequence.  
ACCESSION H55177  
VERSION H55177.1 GI:1108043  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 92)  
Trottauer,J.A., Long,K.R., Murrell,J.R., Stottler,C.J., Gusella,J.F.  
and Buckler,A.J.  
An expression-independent catalog of genes from human chromosome 22  
Genome Res. 5 (3), 214-224 (1995)  
96159527  
On Apr 14, 1993 this sequence version replaced gi:639258.  
CONTACT: Buckler AJ  
Molecular Neurogenetics Unit  
Massachusetts General Hospital  
Building 149, 13th St., Charlestown MA 02129  
Tel: 6177249616  
Fax: 6177265736  
Email: buckler@helix.mgh.harvard.edu  
Insert Length: 2240 Std Error: 0.00  
Seq primer: T3  
High quality sequence stop: 217.  
Location/Qualifiers  
1..92  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="C22\_151"  
/clone\_1ib="Chromosome 22 exon"  
/lab\_host="E. coli DH5a"  
/note="Vector: pBluescriptIIKS+; Site.1: Sal I; Site.2:  
Bam HI (destroyed); Exons were isolated from human  
chromosome 22 specific cosmids using a modification of  
the method of exon amplification (Proc. Natl. Acad. Sci.  
USA 88:4005-4009, 1991). Amplified exons were digested  
with Sal I and Bgl II and subsequently cloned into  
pBluescriptIIKS+ at the Sal I and Bam HI sites."

BASE COUNT 22 a 17 c 24 g 29 t

ORIGIN

Query Match 64.0%; Score 12.8; DB 23; Length 92;  
Best Local Similarity 87.5%; Pred. No. 1.7e+04;  
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ctttaaacatcgcca 19  
||||| ||||| |||  
Db 92 CTTCAAAACATCTCCA 77

RESULT 13  
A1480496 55 bp mRNA EST 09-MAR-1999  
LOCUS vg34a03.x1 Soares\_mammary\_gland\_NBMG Mus musculus cDNA clone  
DEFINITION IMAGE:863212.3 Similar to TR:064306 Q64306 PROLINE-RICH PROTEIN.  
; A1480496  
IMAGE:863212.3  
ACCESSION A1480496  
VERSION A1480496.1 GI:4373722  
KEYWORDS EST.  
SOURCE house mouse.

ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 55)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/cgi/cgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Jun 22, 1998 this sequence version replaced gi:3246745.  
CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
MGI:507300  
This clone was previously sequenced on the 5' end only, this new  
data is from the 3' end  
Trace considered overall poor quality  
Possible reversed clone: similarity on wrong strand  
High quality sequence stop: 1.  
Location/Qualifiers  
1..55  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:863212"  
/clone\_1ib="Soares\_mammary\_gland\_NBMG"  
/sex="male"  
/tissue.type="mammary gland"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Organ: mammary gland; Vector: pT7T3P-Pac  
(Pharmacia) with a modified polylinker; Site.1: Not I;  
Site.2: Eco RI; 1st strand cDNA was primed with a Not I -  
oligo(dT) primer [5',  
TGTACCAATCGAAGTGGAGCGCGCGAATGCTTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT7T3 vector.  
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library.  
constructed and normalized by Bento Soares and M.Fatima  
Bonaldo."

BASE COUNT 17 a 22 c 12 g 3 t 1 others

ORIGIN

Query Match 63.0%; Score 12.6; DB 47; Length 55;  
Best Local Similarity 75.0%; Pred. No. 2.3e+04;  
Matches 15; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ccccttaaacatcgccac 20  
||||| ||||| |||  
Db 7 CCCCTTAAAAAANCCCCGC 26

RESULT 14  
A0014474 84 bp mRNA EST 03-AUG-1998  
LOCUS A0014474 Schizosaccharomyces pombe late log phase cDNA  
DEFINITION Schizosaccharomyces pombe cDNA clone spc09915, mRNA sequence.  
A0014474  
VERSION A0014474.1 GI:3369265  
KEYWORDS EST.  
SOURCE fission yeast.  
ORGANISM Schizosaccharomyces pombe  
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;  
Schizosaccharomycetaceae; Schizosaccharomyces.  
1 (bases 1 to 84)  
Morimyo,M. and Mita,K.  
Identification of expressed sequence tags of Schizosaccharomyces  
pombe  
Unpublished (1998)  
On Jan 9, 1998 this sequence version replaced gi:837494.

**CONTACT:** Mitsunori Morimyo  
Genome Research Group  
National Institute of Radiological Sciences  
9-1, Anagawa-4-chome, Inage-ku, Chiba 263-8555, Japan  
Email: morimyo@nirs.go.jp

**FEATURES**  
source  
location/Qualifiers  
1..84

BASE COUNT ORIGIN	29 a	16 c	9 g	29 t	1 others
----------------------	------	------	-----	------	----------

Query Match	63.0%;	Score 12.6;	DB 41;	Length 84;
Best Local Similarity	75.0%;	Pred. No. 2.2e+04;		
Matches 15; Conservative	0;	Mismatches 5;	Indels 0;	Gaps 0

```

QY      1  cccttaaacatcgccac 20
          | | | | | | | | | |
Db      10  CGCTTTNATACATCGCTAC 29

```

RESULT 15

LOCUS	AU014499	84 bp	mRNA	EST	03-AUG-1998
DEFINITION	AU014499 Schizosaccharomyces pombe late log phase cDNA				
ACCESSION	Schizosaccharomyces pombe cDNA clone spc0947, mRNA sequence.				
VERSION	AU014499				
KEYWORDS	AU014499.1	GI:336290			
SOURCE	EST.				
ORGANISM	fission yeast.				
	Schizosaccharomyces pombe				

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 84)	Morimyo, M. and Mita, K.	Identification of expressed sequence tags of Schizosaccharomyces

JOURNAL  
COMMENT

Unpublished (1998)  
On Jan 9, 1998 this sequence replaced gl:937426

FEATURES	Location/Qualifiers
source	1. .84

BASE COUNT	29 a	16 c	9 g	29 t	1 others
ORIGIN					

Query Match	Score 12.6;	DB 41;	length 84;
Best Local Similarity	75.0%;	Pred. No. 2.2e+04;	
Matches 15; Conservative	0;	Mismatches 5;	Indels 0;
Gaps	0;		
Qy	1 ccccttaaacatgcac	20	
Db	10 GCCTTTNNATCATCGCTAC	29	

Search completed: June 4, 2000, 13:53:22  
Job time: 20946 sec

10/20/2000

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:09:44 ; Search time 244.64 Seconds  
(without alignments)  
10.627 Million cell updates/sec

Title: US-09-164-714-14

Perfect score: 20

Sequence: 1 ccccttaaacatcgccac 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 375880

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/lna/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/lna/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/lna/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/lna/PCFUS.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/lna/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14.2	71.0	50	US-08-171-389-545	Sequence 545, App
2	14.2	71.0	50	US-08-123-936-545	Sequence 545, App
3	14.2	71.0	50	US-08-475-228A-545	Sequence 545, App
4	14.2	71.0	50	US-08-482-080A-545	Sequence 545, App
5	14.2	71.0	50	PCF-US93-12388-545	Sequence 545, App
6	12.6	63.0	22	US-08-117-952-182	Sequence 182, App
7	12.6	63.0	22	US-08-117-952-346	Sequence 346, App
8	12.6	63.0	22	US-08-614-686A-9	Sequence 9, App11
9	12.4	62.0	60	US-08-852-936C-6	Sequence 6, App11
10	12.2	61.0	28	US-08-455-968E-14	Sequence 14, App1
11	12.2	61.0	46	US-08-171-389-308	Sequence 308, App
12	12.2	61.0	46	US-08-123-936-308	Sequence 308, App
13	12.2	61.0	46	US-08-475-228A-308	Sequence 308, App
14	12.2	61.0	46	US-08-482-080A-308	Sequence 308, App
15	12.2	61.0	46	PCF-US93-12388-308	Sequence 308, App
16	12.2	60.0	49	US-08-045-806-22	Sequence 22, App
17	12.2	60.0	49	US-08-366-051B-22	Sequence 22, App1
18	12.2	60.0	50	US-08-171-389-538	Sequence 538, App
19	12.2	60.0	50	US-08-123-936-538	Sequence 538, App
20	12.2	60.0	50	US-08-475-228A-538	Sequence 538, App
21	12.2	60.0	50	US-08-482-080A-538	Sequence 538, App
22	12.2	60.0	50	PCF-US93-12388-538	Sequence 538, App
23	12.2	60.0	66	US-08-852-936C-9	Sequence 9, App11
24	12.2	60.0	80	US-08-276-452A-64	Sequence 64, App1
25	12.2	60.0	80	US-08-798-744-64	Sequence 64, App1
26	11.8	59.0	20	US-08-609-572-6	Sequence 6, App11
27	11.8	59.0	23	US-08-179-557-12	Sequence 12, App1

28	11.8	59.0	89	2	US-08-162-102C-48	Sequence 48, App1
29	11.6	58.0	20	3	US-08-117-952-70	Sequence 70, App1
30	11.6	58.0	21	3	US-07-952-277A-82	Sequence 82, App1
31	11.6	58.0	48	3	US-08-642-045B-21	Sequence 21, App1
32	11.6	58.0	60	1	US-08-478-039-6	Sequence 6, App11
33	11.6	58.0	60	2	US-08-476-349A-6	Sequence 6, App11
34	11.6	58.0	74	4	US-08-708-856A-11	Sequence 11, App1
35	11.6	58.0	74	5	US-09-287-375-11	Sequence 11, App1
36	11.6	58.0	75	5	US-08-828-712-3	Sequence 3, App11
37	11.6	58.0	78	7	5171680-7	Patent No. 5171680
38	11.4	57.0	15	2	US-08-334-847-560	Sequence 560, App
39	11.4	57.0	46	2	US-08-462-388-8	Sequence 8, App1
40	11.4	57.0	48	3	US-08-811-492-124	Sequence 124, App
41	11.2	56.0	34	1	US-08-151-574-37	Sequence 37, App1
42	11.2	56.0	34	1	US-08-146-422-4	Sequence 4, App11
43	11.2	56.0	34	1	US-08-146-424-4	Sequence 4, App11
44	11.2	56.0	34	2	US-08-626-554-20	Sequence 20, App1
45	11.2	56.0	34	3	US-08-419-448-37	Sequence 37, App1

## ALIGNMENTS

RESULT 1  
US-08-171-389-545/C  
Sequence 545, Application US/08171389  
Patent No. 5578444  
GENERAL INFORMATION:  
APPLICANT: Edwards, Cynthia A.  
APPLICANT: Cantor, Charles R.  
APPLICANT: Andrews, Beth M.  
APPLICANT: Turin, Lisa M.  
APPLICANT: Fry, Kirk E.  
TITLE OF INVENTION: Sequence-Directed DNA Binding  
TITLE OF INVENTION: Molecules, Compositions and Methods  
NUMBER OF SEQUENCES: 641  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genelabs Technologies, Inc.  
STREET: 505 Penobscot Drive  
CITY: Redwood City  
STATE: CA  
COUNTRY: USA  
ZIP: 94063  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/171,389  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/123,936  
FILING DATE: 17-SEP-1993.  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/996,783  
FILING DATE: 23-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/723,618  
FILING DATE: 27-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/081,070  
FILING DATE: 22-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0175/G19P3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 545:  
SEQUENCE CHARACTERISTICS:

```

; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Epstein Barr virus R2 (start site
; INDIVIDUAL ISOLATE: 88897)
; US-08-171-389-545

Query Match          71.0%; Score 14.2; DB 1; Length 50;
Best Local Similarity 84.2%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccccttaaacatcgcca 19
    ||||| ||| |||||
Db 28 CCCCTTTAATATCCCA 10

RESULT 2
; US-08-123-936-545/c
; Sequence 545 Application US/08123936
; Patent No. 5726014
; GENERAL INFORMATION:
; APPLICANT: Edwards, Cynthia A.
; APPLICANT: Cantor, Charles R.
; APPLICANT: Andrews, Beth M.
; APPLICANT: Turin, Lisa M.
; TITLE OF INVENTION: Screening Assay for the Detection of
; NUMBER OF SEQUENCES: 640
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/123,936
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/723,618
; FILING DATE: 27-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0075.32/G19P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 545:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Epstein Barr virus R2 (start site
; INDIVIDUAL ISOLATE: 88897)
```

```

US-08-123-936-545

Query Match          71.0%; Score 14.2; DB 2; Length 50;
Best Local Similarity 84.2%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccccttaaacatcgcca 19
    ||||| ||| |||||
Db 28 CCCCTTTAATATCCCA 10

RESULT 3
; US-08-475-228A-545/c
; Sequence 545 Application US/08475228A
; Patent No. 5869241
; GENERAL INFORMATION:
; APPLICANT: Edwards, Cynthia A.
; APPLICANT: Cantor, Charles R.
; APPLICANT: Andrews, Beth M.
; APPLICANT: Turin, Lisa M.
; APPLICANT: Fry, Kirk E.
; TITLE OF INVENTION: Sequence-Directed DNA Binding
; NUMBER OF SEQUENCES: 664
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,228A
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,936
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/723,618
; FILING DATE: 27-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,070
; FILING DATE: 22-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stratford, Carol A.
; REGISTRATION NUMBER: 34,444
; REFERENCE/DOCKET NUMBER: 4600-0175.21/G19P3D2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 545:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Epstein Barr virus R2 (start site
; INDIVIDUAL ISOLATE: 88897)
; US-08-475-228A-545
```

Query Match 71.0%; Score 14.2; DB 3; Length 50;  
Best Local Similarity 84.2%; Pred. No. 49;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccccttaaacatcgcca 19  
||||| ||| ||| |||  
DB 28 CCCCTTTAAATATCCCCA 10

RESULT 4  
US-08-482-080A-545/C  
; Sequence 545, Application US/08482080A  
; Patent No. 6010849  
; GENERAL INFORMATION:  
; APPLICANT: Edwards, Cynthia A.  
; APPLICANT: Cantor, Charles R.  
; APPLICANT: Andrews, Beth M.  
; APPLICANT: Turin, Lisa M.  
; APPLICANT: Fry, Kirk E.  
; TITLE OF INVENTION: Sequence-Directed DNA Binding  
; TITLE OF INVENTION: Molecules, Compositions and Methods  
; NUMBER OF SEQUENCES: 664  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genelabs Technologies, Inc.  
; STREET: 505 Penobscot Drive  
; CITY: Redwood City  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94063  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/482,080A  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/171,389  
; FILING DATE: 20-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/123,936  
; FILING DATE: 17-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/996,783  
; FILING DATE: 23-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/723,618  
; FILING DATE: 27-JUN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/081,070  
; FILING DATE: 22-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brady, John F.  
; REGISTRATION NUMBER: 39,118  
; REFERENCE/DOCKET NUMBER: 4600-0175.20/619P3D1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 324-0980  
; TELEFAX: (650) 324-0960  
; INFORMATION FOR SEQ. ID NO: 545:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 50 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHEICAL: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: Epstein Barr virus R2 (start site  
; INDIVIDUAL ISOLATE: 88897)  
US-08-482-080A-545

Query Match 71.0%; Score 14.2; DB 5; Length 50;  
Best Local Similarity 84.2%; Pred. No. 49;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccccttaaacatcgcca 19  
||||| ||| ||| |||  
DB 28 CCCCTTTAAATATCCCCA 10

RESULT 5  
PCT-US93-12388-545/C  
; Sequence 545, Application PC/TUS9312388  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Sequence-Directed DNA Binding  
; TITLE OF INVENTION: Molecules, Compositions and Methods  
; NUMBER OF SEQUENCES: 641  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genelabs Technologies, Inc.  
; STREET: 505 Penobscot Drive  
; CITY: Redwood City  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94063  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/12388  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/123,936  
; FILING DATE: 17-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/996,783  
; FILING DATE: 23-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fabian, Gary R.  
; REGISTRATION NUMBER: 33,875  
; REFERENCE/DOCKET NUMBER: 4600-0175.41/619PCT2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ. ID NO: 545:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 50 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHEICAL: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: Epstein Barr virus R2 (start site  
; INDIVIDUAL ISOLATE: 88897)  
PCT-US93-12388-545

Query Match 71.0%; Score 14.2; DB 6; Length 50;  
Best Local Similarity 84.2%; Pred. No. 49;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccccttaaacatcgcca 19  
||||| ||| ||| |||  
DB 28 CCCCTTTAAATATCCCCA 10

RESULT 6  
US-08-117-952-182/C  
; Sequence 182, Application US/08117952  
; Patent No. 5851760

```

; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; APPLICANT: Smith, Michael W.
; TITLE OF INVENTION: METHOD FOR GENERATION OF SEQUENCE
; TITLE OF INVENTION: SAMPLED MAPS OF COMPLEX GENOMES
; NUMBER OF SEQUENCES: 797
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/117,952
; FILING DATE: 07-SEP-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/078,471
; FILING DATE: 15-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 182:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Oligonucleotide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
;
US-08-117-952-182

Query Match          63.0%; Score 12.6; DB 3; Length 22;
Best Local Similarity 78.9%; Pred. No. 3e+02;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ccccttaaacatgcga 19
      ||||| ||||| |||
Db 22 CCCCTTCAAAACAGTGCCTA 4

RESULT 7
US-08-117-952-346/c
; Sequence 346, Application US/08117952
; Patent No. 5851760
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; APPLICANT: Smith, Michael W.
; TITLE OF INVENTION: METHOD FOR GENERATION OF SEQUENCE
; TITLE OF INVENTION: SAMPLED MAPS OF COMPLEX GENOMES
; NUMBER OF SEQUENCES: 797
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/117,952
; FILING DATE: 07-SEP-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/078,471
; FILING DATE: 15-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 346:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Oligonucleotide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
;
US-08-117-952-346

Query Match          63.0%; Score 12.6; DB 3; Length 22;
Best Local Similarity 78.9%; Pred. No. 3e+02;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ccccttaaacatgcgcac 20
      ||||| ||||| ||||
Db 21 CTCCTTAGAACATAGCCAC 3

RESULT 8
US-08-614-686A-9
; Sequence 9, Application US/08614686A
; Patent No. 5830692
; GENERAL INFORMATION:
; APPLICANT: B CK, August; MAYER, Dagmar; SCHLENSOG,
; TITLE OF INVENTION: Express System Which Can Be
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Collard & Roe, P.C.
; STREET: 1077 No. 5830692thern Boulevard
; CITY: Roslyn
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 11576
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect Version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/614,686A
; FILING DATE: MARCH 12, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GR19510930
; FILING DATE: 24 MARCH 1995
; APPLICATION NUMBER: GR19514056.7
; FILING DATE: 13 APRIL 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Collard, Allison C.
; REGISTRATION NUMBER: 22,532
; REFERENCE/DOCKET NUMBER: BOCK ET AL.-4
; ATTORNEY/AGENT INFORMATION:
; NAME: Freedman, Edward R.
; REGISTRATION NUMBER: 26,048
```

REFERENCE/DOCKET NUMBER: BOCK ET AL.-4  
ATTORNEY/AGENT INFORMATION:  
NAME: Keusey, Edwin H.  
REGISTRATION NUMBER: 34,361  
REFERENCE/DOCKET NUMBER: BOCK ET AL.-4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (516) 365-9802  
TELEFAX: (516) 365-9805  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 base pairs  
TYPE: nucleotide  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: additional nucleic acid  
DESCRIPTION: synthetic DNA  
IMMEDIATE SOURCE:  
CLONE: Oligo3  
POSITION IN GENOME:  
UNITS: bp  
US-08-614-686A-9

Query Match 63.0%; Score 12.6; DB 3; Length 33;  
Best Local Similarity 78.9%; Pred. No. 3.1e+02;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ccccttaaacatcgcca 19  
||| ||||||| |||  
Db 7 CCCGCCTAACATCTCCA 25

RESULT 9  
US-08-852-936C-6/C  
Sequence 6, Application US/08852936C  
Patent No. 6010878  
GENERAL INFORMATION:  
APPLICANT: DIXIT, VISHVA M.  
APPLICANT: HE, WEI-WU  
APPLICANT: KIRLY, KRISTINE K.  
APPLICANT: RUBEN, STEVEN M.  
TITLE OF INVENTION: INTERLEUKIN-1 BETA CONVERTING  
TITLE OF INVENTION: ENZYME LIKE APOPTOTIC PROTEASE-6  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ratner & Prestia  
STREET: P.O. Box 980  
CITY: Valley Forge  
STATE: PA  
COUNTRY: USA  
ZIP: 19482  
COMPUTER READABLE FORM:  
MEDIUM TYPE: diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/852,936C  
FILING DATE: 08-MAY-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/018,961  
FILING DATE: 05-JUN-1996  
APPLICATION NUMBER: 60/020,344  
FILING DATE: 23-MAY-1996  
APPLICATION NUMBER: 60/017,949  
FILING DATE: 20-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Prestia, Paul F  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: P50483-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-407-0700

TELEFAX: 610-407-0700  
TELEX: 846169  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 60 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-852-936C-6

Query Match 62.0%; Score 12.4; DB 5; Length 60;  
Best Local Similarity 92.9%; Pred. No. 4.2e+02;  
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 ctttaaacatcg 17  
||||||| |  
Db 43 CTTTAAACATCACC 30

RESULT 10  
US-08-455-968E-14  
Sequence 14, Application US/08455968E  
Patent No. 5874283  
GENERAL INFORMATION:  
APPLICANT: Harrington, John L.  
APPLICANT: Hsieh, Chih-Lin  
APPLICANT: Lieber, Michael  
TITLE OF INVENTION: Mammalian Flap-Specific Endonuclease  
NUMBER OF SEQUENCES: 63  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3634  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,968E  
FILING DATE: 30-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 18985-000100  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-576-0200  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (oligonucleotide)  
US-08-455-968E-14

Query Match 61.0%; Score 12.2; DB 3; Length 28;  
Best Local Similarity 82.4%; Pred. No. 4.9e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccccttaaacatcg 17  
||||| ||| |  
Db 9 CCCCTTTAAACTCC 25

```
RESULT 11
US-08-171-389-308
; Sequence 308, Application US/08171389
; Patent No. 5578444
; GENERAL INFORMATION:
; APPLICANT: Edwards, Cynthia A.
; APPLICANT: Cantor, Charles R.
; APPLICANT: Andrews, Beth M.
; APPLICANT: Turin, Lisa M.
; APPLICANT: Fry, Kirk E.
; TITLE OF INVENTION: Sequence-directed DNA Binding
; TITLE OF INVENTION: Molecules, Compositions and Methods
; NUMBER OF SEQUENCES: 641
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/171.389
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123.936
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/723,618
; FILING DATE: 27-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,070
; FILING DATE: 22-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0175/G19P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 308:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Human thrombospondin gene
US-08-171-389-308

Query Match 61.0%; Score 12.2; DB 1; Length 46;
Best Local Similarity 82.4%; Pred. No. 5.2e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
; Patent No. 5726014
; GENERAL INFORMATION:
; APPLICANT: Edwards, Cynthia A.
; APPLICANT: Cantor, Charles R.
; APPLICANT: Andrews, Beth M.
; APPLICANT: Turin, Lisa M.
; APPLICANT: Fry, Kirk E.
; TITLE OF INVENTION: Screening Assay for the Detection of
; TITLE OF INVENTION: DNA-Binding Molecules
; NUMBER OF SEQUENCES: 640
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/123,936
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/723,618
; FILING DATE: 27-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0075.32/G19P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 308:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Human thrombospondin gene
US-08-123-936-308

Query Match 61.0%; Score 12.2; DB 2; Length 46;
Best Local Similarity 82.4%; Pred. No. 5.2e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```

CORRESPONDENCE ADDRESS:
ADDRESS: Genelabs Technologies, Inc.
STREET: 505 Penobscot Drive
CITY: Redwood City
STATE: CA
COUNTRY: USA
ZIP: 94063
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,228A
FILING DATE: 06-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/123,936
FILING DATE: 17-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/996,783
FILING DATE: 23-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/723,618
FILING DATE: 27-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/081,070
FILING DATE: 22-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stratford, Carol A.
REGISTRATION NUMBER: 34,444
REFERENCE/DOCKET NUMBER: 4600-0175.21/619P3D2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 308:
SEQUENCE CHARACTERISTICS:
LENGTH: 46 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Human thrombospondin gene
US-08-475-228A-308

Query Match          61.0%; Score 12.2; DB 3; Length 46;
Best Local Similarity 82.4%; Pred. No. 5.2e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccccttaaacatcgc 17
   |||||||||  |||
Db 17 CCCCTTTAAAGCGCGC 33

RESULT 14
US-08-482-080A-308
; Sequence 308, Application US/08482080A
; Patent No. 6010849
; GENERAL INFORMATION:
; APPLICANT: Edwards, Cynthia A.
; APPLICANT: Cantor, Charles R.
; APPLICANT: Andrews, Beth M.
; APPLICANT: Turin, Lisa M.
; APPLICANT: Fry, Kirk E.
; TITLE OF INVENTION: Sequence-directed DNA Binding
; TITLE OF INVENTION: Molecules, Compositions and Methods
; NUMBER OF SEQUENCES: 664
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
;
```

```

STATE: CA
COUNTRY: USA
ZIP: 94063
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,080A
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/171,389
FILING DATE: 20-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/123,936
FILING DATE: 17-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/996,783
FILING DATE: 23-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/723,618
FILING DATE: 27-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/081,070
FILING DATE: 22-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Brady, John F.
REGISTRATION NUMBER: 39,118
REFERENCE/DOCKET NUMBER: 4600-0175.20/619P3D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 324-0880
TELEFAX: (650) 324-0960
INFORMATION FOR SEQ ID NO: 308:
SEQUENCE CHARACTERISTICS:
LENGTH: 46 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Human thrombospondin gene
US-08-482-080A-308

Query Match          61.0%; Score 12.2; DB 5; Length 46;
Best Local Similarity 82.4%; Pred. No. 5.2e+02;
Matches 14; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccccttaaacatcgc 17
   |||||||||  |||
Db 17 CCCCTTTAAAGCGCGC 33

RESULT 15
PCT-US93-12388-308
; Sequence 308, Application PC/TUS9312388
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Sequence-directed DNA Binding
; TITLE OF INVENTION: Molecules, Compositions and Methods
; NUMBER OF SEQUENCES: 641
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
;
```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/12388  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/123,936  
FILING DATE: 17-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/996,783  
FILING DATE: 23-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0175.41/G19PCT2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 308:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Human thrombospondin gene  
PCT-US93-12388-308

Query Match 61.0%; Score 12.2; DB 6; Length 46;  
Best Local Similarity 82.4%; Pred. No. 5.2e+02;  
Matches 14; Conservative 0; Mismatches 3; Indels 0;

OY 1 ccccttaaacatgc 17  
|||||  
Db 17 CCCCTTTAAAGCGCGC 33

Search completed: June 4, 2000, 16:09:45  
Job time: 28064 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:24:14 ; Search time 322.35 Seconds  
(without alignments)  
34,927 Million cell updates/sec

Title: US-09-164-714-15

Perfect score: 45

Sequence: 1 ggcagcagatgcgaacttaaa.....ctattgcagatcagcttc 45

Scoring table:

IDENTITY\_NDC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 431286

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.4	40.9	51	1	TSSR-1 mutagenic p
2	18	40.0	66	1	Sequence of C. tra
3	18	40.0	90	1	Sequence of Chlamy
4	17.8	39.6	35	1	PCR primer set MPQ
5	17.8	39.6	35	1	Trichosanthin prim
6	17.8	39.6	47	1	PCR primer FL for
7	17.8	39.6	47	1	Light chain primer
8	17.8	39.6	54	1	Sequence encoding
9	17.6	39.1	100	1	Human biallelic po
10	17.4	38.7	83	1	Orphan nuclear hor
11	17.4	38.7	83	1	L57 consensus nucl
12	16.8	37.3	86	1	Oligonucleotide V6
13	16.6	36.9	28	1	Primer used in iso
14	16.6	36.9	47	1	Forward primer for
15	16.6	36.9	47	1	V33995
16	16.6	36.9	80	1	Human gene signatu
17	16.4	36.4	50	1	Human 70 kda heat
18	16.4	36.4	50	1	Human 70 kda heat
19	16.4	36.4	50	1	Test sequence from
20	16.4	36.4	78	1	CANV355 promoter B
21	16.4	36.4	100	1	Candida albicans-s
22	16.2	36.0	69	1	Staphylococcus aur
23	16.2	36.0	81	1	Staphylococcus aur
24	16.2	36.0	83	1	Mouse orphan hormo
25	15.8	35.1	29	1	Mouse VL30 retror
26	15.8	35.1	37	1	Amplification Prim
27	15.8	35.1	38	1	Primer AB140 for I
28	15.8	35.1	66	1	Sequence of C. tra
29	15.8	35.1	66	1	C. trachomatis ser
30	15.6	34.7	38	1	Probe B(1) for bov
31	15.6	34.7	38	1	Probe B(II) for bo
32	15.6	34.7	51	1	5mer synthesised
33	15.6	34.7	88	1	5'-3' oligo 3 for
34	15.6	34.7	90	1	5' leader sequence

c	35	15.4	34.2	41	1	034160	Sequence upstream
c	36	15.4	34.2	41	1	V81550	Oligonucleotide us
c	37	15.4	34.2	44	1	V81549	Oligonucleotide us
c	38	15.4	34.2	50	1	069753	Human cytomegalovir
c	39	15.4	34.2	50	1	T64215	HCWV 2.7 kb transc
c	40	15.4	34.2	50	1	X17503	Test sequence from
c	41	15.4	34.2	51	1	072798	TSSR-1 mutagenic p
c	42	15.4	34.2	87	1	035813	Annealing oligonuc
c	43	15.4	34.2	87	1	035812	Annealing oligonuc
c	44	15.4	34.2	87	1	T04761	PCR primer, HIVTM1
c	45	15.4	34.2	87	1	T04762	PCR primer, HIVTM2

## ALIGNMENTS

RESULT	1	
ID	072792	standard; DNA; 51 BP.
AC	072792;	
DT	08-MAY-1995	(first entry)
DE	TSSR-1 mutagenic primer 3top.	
KW	Kawasaki syndrome; mucocutaneous lymph node syndrome;	
KW	toxic shock syndrome toxin-1; TSSR-1; superantigen; antigen;	
KW	Staphylococcus aureus; T-lymphocyte; immune response; vaccine;	
KW	diagnosis; primer; PCR; polymerase chain reaction; mutagenesis; ss.	
OS	Synthetic.	
PN	W09422474-A.	
PD	13-OCT-1994.	
PF	05-APR-1994; U03719.	
PR	05-APR-1993; US-042876.	
PA	(NADE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.	
PI	Kappler J, Kotzlin B, Leung D, Marrack P;	
DR	WPI: 94-332822/41.	
PT	Treating or preventing Kawasaki Syndrome - by administering a	
PT	superantigen or deriv. to modify a T cell response or provoke an	
PS	immune response	
PS	Disclosure; Fig. 5; 50pp; English.	
CC	The gene for S. aureus TSSR-1 superantigen was isolated by PCR	
CC	amplification of genomic DNA using primers Atop (072788) and Btop	
CC	(072804) and Bbot (072805). Mutant TSSR-1 molecules were	
CC	synthesized using the mutagenic PCR primers given in 072789 and	
CC	072791-803. The method introduced random mutations in approx. 17	
CC	residues in the mature TSSR-1 protein. Selected mutated	
CC	superantigens selectively stimulate only some of the T-cell	
CC	populations stimulated by the wild-type superantigen.	
SQ	Sequence 51 BP; 18 A; 9 C; 9 G; 15 T;	
Query Match	40.9%;	Score 18.4; DB 1; Length 51;
Best Local Similarity	78.6%;	Pred. No. 1.5e+02;
Matches 22; Conservative	0;	Mismatches 6; Indels 0; Gaps 0;
QY	17 ttaaaacattgcagatcagctt 44	
DB	14 TAAAAAACACAGATGCGACATCAGCCT 41	
RESULT	2	
ID	N97082/c	
AC	N97082;	
DT	06-MAR-1992	(first entry)
DE	Sequence of C. trachomatis serovar A major outer membrane protein (MOMP)	
DE	variable domain (VD) gene A-VDIT base pairs 487-552	
KW	Chlamydia trachomatis; antigen; monoclonal antibody; vaccine;	
KW	diagnosis; serotyping; non-immunologic assay; ss.	
OS	Chlamydia trachomatis.	
FT	key	
FT	1. 66	
FT	Location/Qualifiers	
PN	US7324664-A.	
PD	29-AUG-1989.	



```

FT      modified_base      /mod_base= inosine
FT      18                 /tag- c
FT      /mod_base= inosine
FT      modified_base      /tag- d
FT      /mod_base= inosine
PN      US5166056-A.
PD      24-NOV-1992.
PF      04-APR-1989; 333184.
PR      04-APR-1989; US-333184.
PR      07-SEP-1989; US-404326.
PR      09-DEC-1991; US-804293.
PA      (GENE-) GENELABS INC.
PI      Chow TP, Platak M;
PI      WPI; 92-414954/50.
PT      Recombinant tricosanthin protein prodn. in E. coli - for use in
PT      the selective inhibition of viral expression in HIV infected
PT      cells
PS      disclosure: Page 30; 37pp; English.
CC      Primer set MPQ-1 was designed for binding to the anti-sense strand
CC      of the TCS coding region.
CC      The TCS coding sequence was amplified using the primers of Q31828-30.
CC      The amplified prod. has the sequence of Q31827, which was used as a
CC      probe. One isolate, PQ21D, comprises the sequence of Q31826.
CC      The recombinant TCS sequence may be used in the recombinant prodn.
CC      of TCS. TCS can be used for the selective inhibition of viral
CC      expression in HIV-infected human T-cells or macrophages.
SQ      Sequence 35 BP; 8 A; 3 C; 8 G; 5 T;

Query Match      39.6%; Score 17.8; DB 1; Length 35;
Best Local Similarity 57.6%; Pred. No. 2.3e+02;
Matches 19; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY      7 catggcaacttaaacactatggcagatc 39
DB      33 CATNGCCTCTTTRANACRTATTTNGCNCCTC 1

RESULT 6
Q24660 024660 standard; DNA; 47 BP.
AC      024660;
DT      10-NOV-1992 (first entry)
DE      PCR primer FL for CAMPATH-1H light chain.
KM      Polymerase chain reaction; humanised antibody; CAMPATH-1H;
KM      rat anti-human CD18 light chain; YFC51.1.1;
KM      human IgG1 heavy chain; PCR grafting; ss.
OS      Synthetic.
PN      WO9207075-A.
PD      30-APR-1992.
PR      08-OCT-1991; G01744.
PR      10-OCT-1990; GB-022011.
PA      (WELL ) WELLCOME FOUND LTD.
PI      Crowe JS, Lewis AP;
PI      WPI; 92-167155/20.
PT      Prep. of chimeric humanised antibodies - using a new polymerase
PT      chain reaction technique
PS      Example 2; Page 45; 67pp; English.
CC      The YFC51.1.1 rat anti-human -CD18 light chain was humanised as
CC      follows:
CC      Primer EL (Q24659) was used with Primer FL (Q24660) in a PCR
CC      reaction using as template CAMPATH-1H light chain (i.e. humanised
CC      CAMPATH-1 on REI framework; Biotechnology 9:64-68 (1991)) to
CC      produce fragment EFL. Three other PCR reactions were performed on
CC      the same template, generating fragments ABL, CDL and GHL. Fragments
CC      EFL and GHL were combined and used as the template for a PCR
CC      reaction with primers EL and HL (Q24662) to produce fragment EHL.
CC      Similarly, fragment ADL was produced from ABL and CDL using the
CC      primers AL and DL (Q24655 and Q24658, respectively). The products
CC      ADL and EHL were purified and combined in a recombinant PCR
CC      reaction using primers AL and HL. The final humanised light chain

```

```

CC      product, AHL, was cloned into the HindIII site of pUC18 (primers
CC      AL and HL both contain HindIII sites).
SQ      Sequence 47 BP; 9 A; 7 C; 16 G; 15 T;

Query Match      39.6%; Score 17.8; DB 1; Length 47;
Best Local Similarity 75.9%; Pred. No. 2.4e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      15 cttaaaaacactatggcagatcagct 43
DB      11 CTTTCATATACATCTGTGCAGTACTAGCT 39

RESULT 7
Q35185 035185 standard; cDNA; 47 BP.
AC      035185;
DT      08-JUN-1993 (first entry)
DE      Light chain primer FL.
KM      Probe; myeloma; Y3-Ag 1.2.3; primer; rat; monoclonal; antibody; COS;
KM      YFC51.1.1; CD18; humanised; antigen; leukocyte; lung; sepsis; asthma;
KM      endotoxic shock; adult respiratory distress syndrome; inflammation;
KM      immunotoxin; transient expression; PCR; polymerase chain reaction; ss.
OS      Synthetic.
PN      WO9302191-A.
PD      04-FEB-1993.
PF      15-JUL-1992; G01289.
PR      16-JUL-1991; GB-015364.
PA      (WELL ) WELLCOME FOUND LTD.
PI      Crowe S, Sims M, Waldmann H;
PI      WPI; 93-058788/07.
PT      New humanised antibody specific for human CD-18 antigen -
PT      inhibits influx of leukocytes into the lungs, useful for treating
PT      endotoxic shock, adult respiratory distress syndrome, asthma, etc.
PS      Disclosure: Page 48; 59pp; English.
CC      The sequences given in Q35180-87 are primers which were used to
CC      amplify and humanise the light chain isolated from the rat antibody
CC      YFC51.1.1. The light chain of YFC51.1.1 was isolated using a non-
CC      radioactively labelled clone of the light chain from rat myeloma
CC      Y3-Ag 1.2.3. The isolated sequences were amplified, humanised and
CC      constructed into the light chain genes using these primers. The gene
CC      construction, and a corresponding one for the heavy chain (see also
CC      Q35186-95) were transformed into COS cells which transiently
CC      expressed the humanised YFC51.1.1. YFC51.1.1 is a CD18 antibody
CC      which was used as a basis for the production of a humanised antibody
CC      with specificity for CD18 antigen. The antibody may be useful in
CC      treating leukocyte-mediated conditions, such as inhibiting influx of
CC      leukocytes into the lung and other organs during sepsis, endotoxic
CC      shock or adult respiratory distress syndrome. The antibodies may
CC      also be used to treat asthma and inflammation and may form part of
CC      an immunotoxin.
SQ      Sequence 47 BP; 9 A; 7 C; 16 G; 15 T;

Query Match      39.6%; Score 17.8; DB 1; Length 47;
Best Local Similarity 75.9%; Pred. No. 2.4e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      15 cttaaaaacactatggcagatcagct 43
DB      11 CTTTCATATACATCTGTGCAGTACTAGCT 39

RESULT 8
NB1348/C NB1348/C
AC      NB1348;
DT      15-NOV-1990 (first entry)
DE      Sequence encoding modified AA sequence of signal peptide for
DE      egg white lysozyme
KM      Modified egg white lysozyme signal sequence; enzyme; ds.
OS      Homo sapiens.

```

Query Match 39.6%; Score 17.8; DB 1; Length 54;  
Best Local Similarity 62.2%; Pred. No. 2.5e+02;  
Matches 28; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Key Location/Qualifiers  
signal\_peptide 1..54  
/\*tag= a

FT signal\_peptide 1..54  
FT J6333789-A.  
PD 29-SEP-1988.  
PF 23-MAR-1987; 069764.  
PR 23-MAR-1987; JP-069764.  
PA (AUIIN) Tampaku Kagaku Kenk.  
DR WPI; 88-318076/45.  
DR P-PSDB; P81021.  
PT DNA sequence encoding signal peptide -  
PT having good protein secretion activity  
PS Claim 4; Page 523; 11pp; Japanese.  
CC AAs of egg white lysozyme other than 1st, 2nd and 16th-18th residue  
CC are all or mostly substituted with one kind of hydrophobic AA. Using the  
CC modified signal peptide, foreign proteins, esp. human lysozyme, can be  
CC secreted correctly and at a high level.  
SQ Sequence 54 BP; 6 A; 5 C; 16 G; 27 T;

Query Match 39.6%; Score 17.8; DB 1; Length 54;  
Best Local Similarity 62.2%; Pred. No. 2.5e+02;  
Matches 28; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

1 ggaagcattgcaactttaaacattggcagatcatc 45  
47 GCAGCCATATGCACATTAACAAACATATGCACATCAATCTC 3

RESULT 9  
X10693  
ID X10693 standard; DNA; 100 BP.  
AC X10693;  
DE 30-MAR-1999 (first entry)  
KW Human biallelic polymorphic DNA fragment WI-9558.  
KW Polymorphism: biallelic; human; forensic; paternity testing; disease;  
KW detection; phenotypic typing; characteristic; infection; hereditary;  
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;  
KW treatment; marker; ss.  
OS Homo sapiens.  
PN WO9820165-A2.  
PD 14-MAY-1998.  
PF 05-NOV-1997; U20313.  
PR 06-NOV-1996; US-030455.  
PI (MHED) WHITEHEAD INST BIOMEDICAL RES.  
DR WPI; 98-286974/25.  
PT New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease  
PS Claim 1; Page 72; 310pp; English.  
CC X10269-X12937 are human DNA fragments which contain biallelic polymorphic  
CC markers which have been isolated using the primers represented in  
CC X09121-X10268. The base occupying the polymorphic site is indicated by  
CC the appropriate IUPAC-IUB ambiguity code. These fragments can be used in  
CC methods for determining polymorphic forms in an individual for use in  
CC e.g. forensics, paternity testing or for phenotypic typing for diseases  
CC such as agammaglobulinemia, diabetes insipidus, Leech-Nyhan syndrome,  
CC muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial  
CC hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary  
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.  
SQ Sequence 100 BP; 33 A; 24 C; 15 G; 27 T;

Query Match 39.1%; Score 17.6; DB 1; Length 100;  
Best Local Similarity 71.9%; Pred. No. 3.2e+02;  
Matches 23; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

9 tggcaactttaaacattggcagatcatc 40  
24 TGCTATCTTTAGAACACTTTTCAGCAAGATCA 55

RESULT 10  
T14004/c  
ID T14004 standard; CDNA; 83 BP.  
AC T14004;  
DE 19-JUN-1996 (first entry)  
DE Orphan nuclear hormone receptor clone L57 (nt233-315).  
KW Obesity; body weight disorder; diagnosis; therapy; appetite; fat;  
KW gene modulator; adipose; orphan nuclear hormone receptor; ss.  
OS Mus musculus.  
PN WO9605861-A1.  
PD 29-FEB-1996.  
PF 23-AUG-1995; U10918.  
PR 23-AUG-1994; US-294522.  
PR 06-JUN-1995; US-470868.  
PA (MILL-) MILLENIUM PHARM INC.  
PI Tartaglia LA;  
DR WPI; 96-151150/15.  
PT New gene which is differentially expressed in body weight disorders  
PT - corresp. gene prod., antibodies and gene modulators; for  
PT diagnosing and treating a weight disorder, such as obesity  
PS Example 9; Fig 13; 162pp; English.  
CC A set point paradigm was used to identify mouse gene sequences  
CC which are differentially expressed and which may contribute to body  
CC wt. disorders and/or may be involved in body wt. regulation or  
CC appetite modulation. cDNA clone L57 was isolated. Nucleotides  
CC 39-150 (T14002), 151-269 (T14003), 233-315 (T14004) and 4-41  
CC (T14005) correspond to nucleotides 1321-1432, 1432-1550, 1513-1595  
CC and 1285-1322, respectively, of the mouse orphan nuclear hormone  
CC receptor. L57 gene expression is higher in the liver of  
CC underweight mice than in that of overweight littermates. L57,  
CC and other identified genes (see T13979-T14001 and T14006-T14008),  
CC can be used to diagnose body wt. disorders.  
SQ Sequence 83 BP; 29 A; 21 C; 8 G; 24 T;

Query Match 38.7%; Score 17.4; DB 1; Length 83;  
Best Local Similarity 64.9%; Pred. No. 3.7e+02;  
Matches 24; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

2 gacgcattgcaactttaaacattggcagat 38  
61 GAAGAYCTGCAACTTTAGAACATATTGTGTAATAAT 25

RESULT 11  
X03753/c  
ID X03753 standard; DNA; 83 BP.  
AC X03753;  
DE 31-MAR-1999 (first entry)  
DE L57 consensus nucleotide sequence #3.  
KW Body weight disorder; obesity; appetite regulation; thermoregulation;  
KW anorexia; abnormal food intake; cachexia; thermogenesis; ss.  
OS Synthetic.  
PN US5853975-A.  
PD 29-DEC-1998.  
PF 26-FEB-1997; 807861.  
PR 26-FEB-1997; US-807861.  
PR 23-AUG-1994; US-294522.  
PR 06-JUN-1995; US-470868.  
PR 23-AUG-1995; US-518878.  
PA (MILL-) MILLENIUM PHARM INC.  
PI Tartaglia LA;  
DR WPI; 99-094892/08.  
PT Identifying modulators of c5 protein uncoupling activity - used as

Query Match	Best Local Similarity	Score	DB 1;	Length	83;
Matches	24;	Conservative	1;	Mismatches	12;
				Indels	0;
				Gaps	0
Qy	2	gagccatgacacattaaacacattggcagcat	38		
Db	61	GAAGAYCTGCGCAACTTACGACGTAATTTGTAAATAAT	25		
RESULT	12				
ID	V16538	standard; DNA; 86 BP.			
AC	V16538;				
DT	11-JUN-1998	(first entry)			
DE	Oligonucleotide V6635.				
KW	Nucleic acid detection; Identification;				
MM	variant diagnostic base sequence; ss.				
OS	Synthetic.				
PN	GB2312747-A.				
PD	05-NOV-1997.				
PF	29-APR-1997: 008581.				
PR	04-MAY-1996: GB-009441.				
PI	(ZENNE) ZENNECA LTD.				
PI	Brownie J, Little S, Whitcombe DM;				
DR	WPI; 97-506147/47.				
PT	Detection of a diagnostic base sequence in a nucleic acid sample -				
PT	using primers with non-complementary tails, is useful for e.g.				
PT	diagnosis of diseases and identification of samples in forensic				
PT	medicine				
PS	Disclosure; Page 24; 57pp; English.				
CC	The present oligonucleotide was used to exemplify the method of the				
CC	specification, which describes the detection of a diagnostic nucleic				
CC	acid base. The method comprises contacting the sample with a diagnostic				
CC	primer for the diagnostic base under hybridising conditions and in the				
CC	presence of appropriate nucleoside triphosphates and an agent for				
CC	polymerisation. The diagnostic primer has a non-complementary tail				
CC	sequence comprising a tag and a detector region. Any extension products				
CC	are used as a template for extension of a further primer which hybridises				
CC	to a locus at a distance from the diagnostic base. The sample is				
CC	contacted with a tag primer which selectively hybridises to the				
CC	complement of the tag sequence in an extension product of the further				
CC	primer. The presence or absence of the diagnostic base is detected by				
CC	reference to the detector region in the further primer extension product.				
CC	The method is used to identify one or more variant diagnostic base				
CC	sequences against a background of normal diagnostic base sequences.				
CC	Detection of nucleic acids is important for detection and diagnosis of				
CC	allies or DNA sequences associated or linked to genes responsible for				
CC	genetic diseases in humans and other species, detection and diagnosis of				
CC	neoplasms, detection and distinction between different pathogens,				
CC	determining the purity of animal strains and pedigrees, and				
CC	distinguishing and identifying different human and animal samples in				
CC	forensic medicine.				
Sequence	86 BP;	22 A;	22 C;	25 G;	17 T;

```

Query Match          37.3%: Score 16.8; DB 1; Length 86;
Best Local Similarity 75.0%: Pred. No. 6.3e+02;
Matches 21; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY      2 gaccatggcgaacttaaacactat 29
        ||||| ||||| ||||| |||||
Db      53 GACGGCTGCGCACCATTTAAGAAATAAT 80

RESULT 13
T88000/c
ID      T88000 standard; DNA: 28 BP.
AC      T88000;
DE      27-APR-1998 (first entry)
DT      Primer used in isolation of Werner's syndrome 3 gene.
KW      Werner's syndrome 3; WS-3; diagnosis; probe; examination;
KW      prevention; antibody; reagent; human ontogenesis; detection;
KW      PCR primer; ss.
OS      Synthetic.
PN      J09238683-A.
PD      16-SEP-1997.
PE      12-MAR-1996; 055144.
PR      12-MAR-1996; JP-055144.
PA      (EITJ-) EJIN KENKYUSHO KK.
        WPI: 97-506552/47.
PT      Werner's syndrome 3 gene - useful to develop products to detect,
PT      diagnose and study the disease
PS      Example 4: Page 22; 28pp; Japanese.
CC      The present sequence was used in the isolation of the Werner's
CC      syndrome 3 (WS-3) gene, which can be used as a diagnostic probe for
CC      WS related diseases. It may also be used to polyclonal antibody
CC      such diseases. The protein or a monoclonal or polyclonal antibody
CC      raised against it can be used as a reagent for studies on human
CC      ontogenesis, or WS-3 detection. An oligonucleotide probe specific
CC      to the WS-3 gene can be used as a reagent for the detection of the
CC      WS-3 gene.
SQ      Sequence 28 BP: 7 A; 4 C; 8 G; 9 T;

Query Match          36.9%: Score 16.6; DB 1; Length 28;
Best Local Similarity 82.6%: Pred. No. 6.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      10 ggcacctaataaacactatgg 32
        | ||||| || ||| |||||
Db      28 GCCACTTCACACACATTAATTGG 6

RESULT 14
V33999/c
ID      V33999 standard; DNA: 47 BP.
AC      V33999;
DE      25-JAN-1999 (first entry)
DT      Forward primer for truncated apopain p12 subunit gene.
KW      inhibitor; apopain; binding site; crystal structure; primer; PCR;
KW      amplification; Escherichia coli; apoptosis; ss.
OS      Synthetic.
PN      US5634228-A.
PD      10-NOV-1998.
PE      13-FEB-1997; 800007.
PR      13-FEB-1997; US-800007.
PA      (MERI ) MERCK & CO INC.
PI      (MERI ) MERCK FROST CANADA INC.
PI      Becker JW, Fazil KM, Gallant M, Gareau Y, Labelle M,
PI      Nicholson DM, Peterson EP, Rasper DM, Rotonda J,
PI      Fuiel R, Thornderry NA, Vaillancourt JP;
        WPI: 99-008706/01.
PT      Identifying inhibitors of apopain by rational drug design - from
PT      ability to bind to the enzyme's substrate binding domain,
PT      potentially useful for inhibiting apoptosis; e.g in treatment of
PT      acquired immune deficiency syndrome, autoimmune disease, infections

```

```
Search completed: June  4, 2000, 16:24:15
Job time: 28911 sec
```

PT etc  
PS Example 1: Column 21; 22pp; English.  
CC The invention relates to the identification of inhibitors of apopain  
CC by designing and synthesising a potential inhibitor that will form  
CC non-covalent bonds with amino acids in the apopain substrate binding  
CC site, based upon the crystal structure co-ordinates of an  
CC apopain:acetyl-DEVD-CHO complex, testing it for apopain inhibition.  
CC Production of apopain for the binding studies comprises expression of  
CC the p17 and p12 subunits separately. The primers V33999-V34000 were  
CC used to generate a truncated p12 subunit comprising amino acid residues  
CC Ser310-His402 and truncated for the binding studies. The protein was  
CC expressed in *E. coli*. The apopain inhibitors are potentially useful  
CC for inhibiting apoptosis.  
SQ Sequence 47 BP; 9 A; 8 C; 16 G; 14 T;

Query Match	36.98;	Score 16.6;	DB 1;	Length 47;
Best Local Similarity	71.0%;	Pred. No. 6.8e+02;		
Matches 22;	Conservative 0;	Mismatches 9;	Indels 0;	Gaps 0

```

QY      3  acgccatgcaactttaaaacactattgac 33
          |||||  |||  |||  |||  |||  |||
Db      44  ACGCCATGTCTATCATCAACACCACTCATGAC 144

```

RESULT 15

ID V33995/c

AC V33995.

AC V33995 standard; DNA; 47 BP.

DT 25-JAN-1999 (first entry)

DE Forward primer for apopain p12 subunit gene.

KW Inhibitor; apopain; binding site; crystal structure; primer; PCR;

KW amplification; Escherichia coli; apoptosis; ss.

OS Synthetic.

PN US5834228-A.

PD 10-NOV-1998.

PF 13-FEB-1997; 800007.

PR 13-FEB-1997; US-800007.

PA (MERI ) MERCK & CO INC.

PA Becker JW, Fazil KM, Gallant M, Gareau Y, Labelle M,

PI Nicholson DW, Peterson EP, Rasper DM, Rotonda J,

PI Ruel R, Thorntberry NA, Vaillancourt JP;

DR WPI: 99-008706/01.

PT Identifying inhibitors of apopain by rational drug design - from

PT ability to bind to the enzyme's substrate binding domain,

PT potentially useful for inhibiting apoptosis, e.g in treatment of

PT acquired Immune deficiency syndrome, autoimmune disease, infections

PT etc

PS Disclosure; Column 7; 22pp; English.

CC The invention relates to the identification of inhibitors of apopain

CC by designing and synthesising a potential inhibitor that will form

CC non-covalent bonds with amino acids in the apopain substrate binding

CC site, based upon the crystal structure co-ordinates of an

CC apopain:acetyl-DEVD-CHO complex, testing it for apopain inhibition.

CC Production of apopain for the binding studies comprises expression of

CC the p17 and p12 subunits separately. The primers V33995-V33996 were

CC used to amplify the coding sequence for the p12 subunit for the binding

CC studies. The protein was expressed in E. coli. The apopain inhibitors

CC are potentially useful for inhibiting apoptosis.

CC Sequence 47 BP; 9 A; 8 C; 14 G; 14 T;

Query Match	36.9%	Score 16.6;	DB 1;	Length 47;
Best Local Similarity	71.0%;	Pred. No. 6.8e+02;		
Matches 22; Conservative	0;	Mismatches 9;	Indels 0;	Gaps 0

```

Oy      3  acgccatgycaacctttaaaacactattgc 33
          |||||  |||  |||  |||  |||  |||
Db      44  ACGCCATGTCATCATCAACACCACTCATGAC 14

```

Tue Jun 6 08:33:25 2000

us-09-164-714-15.rng

---

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 13:53:22 ; Search time 4521.53 Seconds  
(without alignments)  
40.339 Million cell updates/sec

Title: US-09-164-714-15

Perfect score: 45

Sequence: 1 ggaagccatggcaactttaa.....ctattggcagatcacgttc 45

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 202661650 residues

Total number of hits satisfying chosen parameters: 119968

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: gb\_est1: \*  
21: gb\_est2: \*  
22: gb\_est3: \*  
23: gb\_est4: \*  
24: gb\_est5: \*  
25: gb\_est6: \*  
26: gb\_est7: \*  
27: gb\_est8: \*  
28: gb\_est9: \*  
29: gb\_est10: \*  
30: gb\_est11: \*  
31: gb\_est12: \*  
32: gb\_est13: \*  
33: gb\_est14: \*  
34: gb\_est15: \*  
35: gb\_est16: \*  
36: gb\_est17: \*  
37: gb\_est18: \*  
38: gb\_est19: \*  
39: gb\_est20: \*  
40: gb\_est21: \*  
41: gb\_est22: \*  
42: gb\_est23: \*  
43: gb\_est24: \*  
44: gb\_est25: \*

45: gb\_est26: \*  
46: gb\_est27: \*  
47: gb\_est28: \*  
48: gb\_est29: \*  
49: gb\_est30: \*  
50: gb\_est31: \*  
51: gb\_est32: \*  
52: em\_est20: \*  
53: em\_est21: \*  
54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*  
59: gb\_est33: \*  
60: gb\_est34: \*  
61: gb\_est35: \*  
62: gb\_est36: \*  
63: gb\_est37: \*  
64: gb\_est38: \*  
65: em\_est27: \*  
66: em\_est28: \*  
67: em\_est29: \*  
68: em\_est30: \*  
69: gb\_est39: \*  
70: gb\_est40: \*  
71: gb\_est41: \*  
72: gb\_est42: \*  
73: gb\_est43: \*  
74: gb\_est44: \*  
75: em\_est31: \*  
76: em\_est32: \*  
77: em\_est33: \*  
78: em\_est34: \*  
79: gb\_est45: \*  
80: gb\_est46: \*  
81: gb\_est47: \*  
82: gb\_est48: \*  
83: gb\_est49: \*  
84: gb\_est50: \*  
85: gb\_est51: \*  
86: em\_est52: \*  
87: em\_est53: \*  
88: em\_est54: \*  
89: gb\_est55: \*  
90: gb\_est56: \*  
91: gb\_est57: \*  
92: gb\_est58: \*  
93: gb\_est59: \*  
94: gb\_est60: \*  
95: em\_est61: \*  
96: em\_est62: \*  
97: em\_est63: \*  
98: em\_est64: \*  
99: em\_est65: \*  
100: em\_est66: \*  
101: em\_est67: \*  
102: gb\_est68: \*  
103: gb\_est69: \*  
104: em\_est70: \*  
105: gb\_est71: \*  
106: gb\_est72: \*  
107: gb\_est73: \*  
108: gb\_est74: \*  
109: gb\_est75: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result

Query

SUMMARIES



QY 5 gccatggcaactttaaacacatattggcagtcac 39  
||||| ||||| ||||| ||| ||| |||  
Db 46 GCCATTGCACAGCAAAAACGTCATCGAATAC 80

RESULT 3  
A1152611/c  
LOCUS A1152611 97 bp mRNA EST 30-SEP-1998  
DEFINITION udb89e04.r1 Soares\_NMPu Mus musculus cDNA clone IMAGE:1478046 5'  
similar to TR:Q14731 Q14731 LAMININ AH ; mRNA sequence.  
ACCESSION A1152611  
VERSION A1152611.1 GI:3681080  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 97)  
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, R., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMT Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Apr 14, 1993 this sequence version replaced gi:639170.  
CONTACT: Marra M/Mouse EST Project  
WashU-HMT Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:926402  
Trace considered overall poor quality  
Possible reversed clone: similarity on wrong strand  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers  
1. 97  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1478046"  
/clone\_lib="Soares\_NMPu"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; 1st strand cDNA was prepared from  
pregnant mouse uterus, and was then primed with a Not I -  
oligo(dT) primer. Double-stranded cDNA was ligated to Eco  
RI adaptors (Pharmacia), digested with Not I and cloned  
into the Not I and Eco RI sites of the modified pT73  
vector. Library is normalized. Library was constructed by  
Bento Soares and M. Fatima Bonaldo."

BASE COUNT 25 a 19 c 21 g 32 t  
ORIGIN

Query Match 41.3%; Score 18.6; DB 42; Length 97;  
Best Local Similarity 65.9%; Pred. No. 4.2e+03;  
Matches 27; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 3 agccatggcaactttaaacacatattggcagtcac 43  
||||| ||||| ||||| ||| ||| |||  
Db 77 AGCGAATTTCAATTTTAACCATATTGAAGACTCATCT 37

RESULT 4  
AA700885/c

LOCUS AA700885 61 bp mRNA EST 19-DEC-1997  
DEFINITION zj40a01.s1 Soares\_fetal\_liver\_spleen\_INFLS\_S1 Homo sapiens cDNA  
clone IMAGE:452712 3' similar to SW:ID: HUMAN Q02363 DNA-BINDING  
PROTEIN INHIBITOR ID-2 ; mRNA sequence.  
ACCESSION AA700885  
VERSION AA700885.1 GI:2704050  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 61)  
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisler, G., Jost, S.,  
Kizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, R., Steptoe, M., Tan, F.,  
Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397509.  
CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Trace considered overall poor quality  
Possible reversed clone: similarity on wrong strand  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers  
1. 61  
/organism="Homo sapiens"  
/db\_xref="GDB:1389068"  
/db\_xref="taxon:9606"  
/clone="IMAGE:452712"  
/clone\_lib="Soares\_fetal\_liver\_spleen\_INFLS\_S1"  
/sex="male"  
/dev\_stage="20 week post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)  
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;  
this is a subtracted version of the original Soares fetal  
liver spleen INFLS library. 1st strand cDNA was primed  
with a Pac I - oligo(dT) primer [5',  
ACCTGGAGAAATTAATTAAGATCTTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 9 a 9 c 21 g 22 t  
ORIGIN

Query Match 39.1%; Score 17.6; DB 37; Length 61;  
Best Local Similarity 83.3%; Pred. No. 1e+04;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 12 caacttaaaaacacatattggcag 35  
||||| ||||| ||||| ||| ||| |||  
Db 56 CAACTTAACACAGATATTGTCAG 33

RESULT 5  
R86574  
LOCUS R86574 92 bp mRNA EST 17-APR-1995  
DEFINITION RABES149M Rabbit Osteoclast, Dennis Sakai Oryctolagus cuniculus  
cDNA clone PRABOC149 3', mRNA sequence.  
ACCESSION R86574  
VERSION R86574.1 GI:947228  
KEYWORDS EST.

SOURCE  
ORGANISM Oryctolagus cuniculus.  
Oryctolagus cuniculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
REFERENCE 1 (bases 1 to 92)  
AUTHORS Sakai, D., Tong, H.-S. and Minkin, C.  
TITLE Osteoclast Molecular Phenotyping by Random cDNA Sequencing  
JOURNAL Bone 17 (2), 111-119 (1995)  
MEDLINE 96021365  
COMMENT On May 8, 1995 this sequence version replaced gi:800292.  
Other ESTs: RABEST149T  
Contact: Sakai D  
Basic Sciences  
University of Southern California  
USC School of Dentistry, 925 West 34th Street, DEN-4220, Los  
Angeles, CA 90089-0641  
Tel: 2137405563  
Fax: 2137407560  
Email: sakai@molbio.usc.edu  
Insert Length: 465 Std Error: 0.00  
Seq primer: M13 forward  
High quality sequence stop: 320.  
Location/Qualifiers

FEATURES  
source  
1..92  
/organism="Oryctolagus cuniculus"  
/strain="New Zealand White"  
/db\_xref="taxon:9986"  
/clone\_lib="PRABOC149"  
/lab\_host="E. coli DH12S"  
/note="Vector: pSPORT1; Site\_1: Sali; Site\_2: NotI;  
Poly(A)+ RNA was purified from a 97% pure population of  
osteoclasts prepared from the long bones of 10 day old  
rabbits. First strand cDNA was synthesized by priming  
with an oligo(dT)-NotI anchor-primer and second strand  
cDNA was synthesized by replacement synthesis as described  
by Gubler and Hoffman (Gene 25:283, 1983). Following the  
addition of Sali adapters and NotI digestion, the cDNA was  
cloned between the Sali (50) and NotI (30) sites of the  
pSPORT1 (BRL) plasmid vector."

BASE COUNT  
ORIGIN 24 a 23 c 21 g 22 t 2 others

Query Match 39.1%; Score 17.6; DB 23; Length 92;  
Best Local Similarity 65.0%; Pred. No. 9.8e+03;  
Matches 26; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 3 acgcacatggcaacttaaaacacattggcagatcagc 42  
1 ||||| ||||| ||||| ||||| |||||  
Db 3 AAGCATGATCTACTTATTAACACGGTGGGATTCTCAAC 42

RESULT 6  
D20993 85 bp mRNA EST 30-JUL-1996  
LOCUS HUMGS01975 Human promyelocyte Homo sapiens cDNA clone mp2228 3',  
DEFINITION mRNA sequence.  
ACCESSION D20993  
VERSION D20993.1 GI:504813  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 85)  
AUTHORS Okubo, K., Fukushima, A., Yoshii, J., Niiyama, T., Kojima, Y.,  
Yoshinari, H., Arimoto, J., and Matsubara, K.  
TITLE Gene expression of human promyelocytic cell line HL60 before and  
after induction of differentiation. A new application of 3'directed  
cDNA sequencing  
JOURNAL Unpublished (1993)  
COMMENT Contact: Okubo, K., Fukushima, A., Yoshii, J., Niiyama, T., Kojima, Y.,

FEATURES  
source  
1..85  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="mp2228"  
/clone\_lib="Human promyelocyte"  
/note="Female, adult, cell\_line = HL60, cell\_type =  
promyelocyte."

BASE COUNT  
ORIGIN 27 a 18 c 15 g 22 t 3 others

Query Match 38.7%; Score 17.4; DB 20; Length 85;  
Best Local Similarity 68.6%; Pred. No. 1.2e+04;  
Matches 24; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 1 ggaacgcacatggcaacttaaaacacattggcagc 35  
1 || ||| ||||| | ||||| |||  
Db 11 GGAACACAGCGCACTAGTATCTACTATTAGCAG 45

RESULT 7  
A1130336 97 bp mRNA EST 26-MAR-1999  
LOCUS A1130336  
DEFINITION SMOVL3CANI7F11 Onchocerca volvulus infective larva cDNA  
(SMA94WL-OVL3) Onchocerca volvulus cDNA clone onchl126 5' similar  
to WP:044E4.4 CE08718 RNA-BINDING PROTEIN ; mRNA sequence.  
ACCESSION A1130336  
VERSION A1130336.1 GI:3600354  
KEYWORDS EST.  
SOURCE Onchocerca volvulus.  
ORGANISM Onchocerca volvulus.  
Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;  
Filarioidea; Onchocercidae; Onchocerca.  
REFERENCE 1 (bases 1 to 97)  
AUTHORS Williams, S.A., Lizotte-Maniowski, M., Laney, S., Wenhong, L.,  
Hillier, L., Allen, M., Bowles, L., Geisel, S., Jost, S., Kucaba, T.,  
Martin, J., Steptoe, M., Theising, B., White, Y., Wylie, T.,  
Chappell, J., Person, B., Gibbons, M., Harvey, N., Pape, D.,  
Chamberlain, A., Morales, R., Schurk, R., Rittler, E., Kohn, S.,  
Underwood, K. and Marra, M.  
Molecular Parasitology OVL3  
Unpublished (1998)  
On Oct 30, 1996 this sequence version replaced gi:1656870.  
Contact: Steven A. Williams  
Molecular Parasitology  
Smith College Department of Biological Sciences  
Department of Biological Sciences, Clark Science Center, Smith  
College, Northampton, MA, 01063, USA  
Tel: 4135853826  
Fax: 4135853786  
Email: genome@smith.edu  
The library was constructed by Wenhong Lu. The library is available  
from Dr. S.A. Williams, email genome@smith.edu when requesting this  
clone from Dr. Williams, please reference the Williams lab clone id  
- SMOVL3CANI7F11  
Trace considered overall poor quality  
Possible reversed clone: similarity on wrong strand  
Seq primer: r3 ET from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers

FEATURES  
source  
1..97  
/organism="Onchocerca volvulus"  
/strain="Siera Leone"  
/db\_xref="taxon:6282"  
/clone\_lib="onchl126"  
/clone\_lib="Onchocerca volvulus infective larva cDNA  
(SMA94WL-OVL3)"  
/lab\_host="XL1-Blue MRF"











---

1000 1000 1000 1000 1000 1000 1000 1000 1000 1000

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:09:45 ; Search time 244.64 Seconds  
(without alignments)  
23.910 Million cell updates/sec

Title: US-09-164-714-15

Perfect score: 45  
Sequence: 1 ggaagccatgcaacttaaa.....ctatggcagtcacgttc 45

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 375880

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/5C.COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/5D.COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/6.COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/PCBUS.COMB.seq:\*  
7: /cgn2\_6/prodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.4	40.9	45	6	PCT-US95-14639-6
2	17.8	39.6	47	3	US-08-039-198B-18
3	17.8	39.6	47	4	US-08-182-067-23
4	17.8	39.6	47	4	US-08-465-313-23
5	17.4	38.7	83	2	US-08-518-878B-47
6	17.4	38.7	83	2	US-08-294-522B-47
7	17.4	38.7	83	3	US-08-807-861A-47
8	17.4	38.7	83	3	US-08-470-868A-47
9	16.6	36.9	47	3	US-08-800-007A-3
10	16.6	36.9	47	3	US-08-800-007A-7
11	16.4	36.4	50	1	US-08-171-389-382
12	16.4	36.4	50	2	US-08-123-936-382
13	16.4	36.4	50	3	US-08-475-228A-382
14	16.4	36.4	50	6	PCT-US95-12388-382
15	16.4	36.4	50	5	PCT-US95-12388-382
16	16.4	36.4	100	1	US-08-145-705A-2
17	16.2	36.0	83	2	US-08-518-878B-32
18	16.2	36.0	83	2	US-08-294-522B-32
19	16.2	36.0	83	3	US-08-807-861A-32
20	16.2	36.0	83	3	US-08-470-868A-32
21	15.8	35.1	29	4	US-08-600-999-13
22	15.8	35.1	37	2	US-08-616-998-2
23	15.6	34.7	51	3	US-08-636-113-1
24	15.4	34.2	41	5	US-09-109-063-45
25	15.4	34.2	44	5	US-09-109-063-44
26	15.4	34.2	44	5	US-09-109-063-44
27	15.4	34.2	45	6	PCT-US95-14639-12

28	15.4	34.2	50	1	US-08-171-389-503	Sequence 503, App
29	15.4	34.2	50	2	US-08-123-936-503	Sequence 503, App
30	15.4	34.2	50	3	US-08-475-228A-503	Sequence 503, App
31	15.4	34.2	50	5	US-08-482-080A-503	Sequence 503, App
32	15.4	34.2	50	6	PCT-US93-12388-503	Sequence 503, App
33	15.4	34.2	77	1	US-08-117-374-14	Sequence 14, Appl
34	15.4	34.2	77	1	US-08-280-263-14	Sequence 14, Appl
35	15.4	34.2	77	6	PCT-US94-10256-14	Sequence 14, Appl
36	15.4	34.2	87	1	US-08-105-483-403	Sequence 403, App
37	15.4	34.2	87	1	US-08-105-483-403	Sequence 403, App
38	15.4	34.2	87	2	US-08-709-209-404	Sequence 404, App
39	15.4	34.2	87	2	US-08-709-209-404	Sequence 404, App
40	15.4	34.2	87	2	US-08-303-275-107	Sequence 107, App
41	15.4	34.2	87	2	US-08-303-275-108	Sequence 108, App
42	15.4	34.2	87	2	US-08-458-101-403	Sequence 403, App
43	15.4	34.2	87	2	US-08-458-101-404	Sequence 404, App
44	15.4	34.2	87	3	US-08-417-210A-129	Sequence 129, App
45	15.4	34.2	87	3	US-08-417-210A-130	Sequence 130, App

#### ALIGNMENTS

RESULT 1

PCT-US95-14639-6

Sequence 6, Application PC/TUS9514639

GENERAL INFORMATION:

APPLICANT: PHILIPA MARRACK

APPLICANT: JOHN KAPLER

APPLICANT: RICHARD SHMONKEVITZ

APPLICANT: MASAZUMI MATSUMURA

TITLE OF INVENTION: PROTECTIVE EFFECTS OF MUTATED

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Swanson & Bratschun, L.L.C.

STREET: 8400 E. Prentice Ave., Suite 200

CITY: Englewood

STATE: Colorado

COUNTRY: USA

ZIP: 80111

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch

COMPUTER: IBM pc compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: WordPerfect 6.0 for windows

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/14639

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/338,373

FILING DATE: 14-November-1994

ATTORNEY/AGENT INFORMATION:

NAME: Barry J. Swanson

REGISTRATION NUMBER: 33,215

REFERENCE/DOCKET NUMBER: SUP020/PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 793-3333

TELEFAX: (303) 793-3433

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 45 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

PCT-US95-14639-6

Query Match 40.9%; Score 18.4; DB 6; Length 45;  
Best Local Similarity 78.6%; Pred. No. 47;  
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;



FILING DATE: 05-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/182,067  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CB92/01289  
FILING DATE: 15-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9115364.3  
FILING DATE: 16-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BROOK, DAVID E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: LYNX91-01A2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 47 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHEICAL: NO  
ANTI-SENSE: YES  
US-08-465-313-23

Query Match 39.6%; Score 17.8; DB 4; Length 47;  
Best Local Similarity 75.9%; Pred. No. 82;

Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 15 ctttaaacactatgagcatgacgt 43  
||| ||| ||| ||| ||| ||| ||| |||  
DB 11 CTTTCAATACTGTGGCAGTAGTAGT 39

RESULT 5  
US-08-518-878B-47/c  
; Sequence 47, Application US/08518878B  
; Patent No. 5702902  
; GENERAL INFORMATION:  
; APPLICANT: Tartaglia, Louis A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TREATMENT OF BODY WEIGHT DISORDERS, INCLUDING OBESITY  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/518,878B  
; FILING DATE: 23-AUG-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-036  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 47:

SEQUENCE CHARACTERISTICS:  
LENGTH: 83 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-518-878B-47

Query Match 38.7%; Score 17.4; DB 2; Length 83;  
Best Local Similarity 64.9%; Pred. No. 1.3e+02;  
Matches 24; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

OY 2 gacgcacatgcaactttaaacactatgacgat 38  
||| : ||| ||| ||| ||| ||| ||| |||  
DB 61 GAAGATCTGGCACTTAGAAGTAATTTGTAAAT 25

RESULT 6  
US-08-294-522B-47/c  
; Sequence 47, Application US/08294522B  
; Patent No. 5741666  
; GENERAL INFORMATION:  
; APPLICANT: Tartaglia, Louis A.  
; TITLE OF INVENTION: Compositions and Methods for the  
; TREATMENT OF BODY WEIGHT DISORDERS, INCLUDING OBESITY  
; NUMBER OF SEQUENCES: 48  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/294,522B  
; FILING DATE: 23-AUG-1994  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-015  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864/9741  
; INFORMATION FOR SEQ ID NO: 47:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 83 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
US-08-294-522B-47

Query Match 38.7%; Score 17.4; DB 2; Length 83;  
Best Local Similarity 64.9%; Pred. No. 1.3e+02;  
Matches 24; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

OY 2 gacgcacatgcaactttaaacactatgacgat 38  
||| : ||| ||| ||| ||| ||| ||| |||  
DB 61 GAAGATCTGGCACTTAGAAGTAATTTGTAAAT 25

RESULT 7  
US-08-807-861A-47/c  
; Sequence 47, Application US/08807861A  
; Patent No. 5853975



```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732-594-6734
; TELEFAX: 732-594-4720
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 47 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-800-007A-3

Query Match          36.9%; Score 16.6; DB 3; Length 47;
Best Local Similarity 71.0%; Pred. No. 2.4e+02;
Matches 22; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 accgcagcgcaacttaaacacttggc 33
Db 44 ACCGCATGTCATCATCAACACCACTCATGAC 14

RESULT 10
US-08-800-007A-7/c
; Sequence 7, Application US/08800007A
; Patent No. 5834228
; GENERAL INFORMATION:
; APPLICANT: Becker, Joseph
; APPLICANT: Nicholson, Donald
; APPLICANT: Rotonda, Jennifer
; APPLICANT: Thorderry, Nancy
; APPLICANT: Fazil, Kimberly
; APPLICANT: Gallant, Michel
; APPLICANT: Gareau, Yves
; APPLICANT: Labelle, Marc
; APPLICANT: Peterson, Erin
; APPLICANT: Rasper, Dita
; TITLE OF INVENTION: CRYSTAL STRUCTURE OF APOPAIN
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JOSEPH A. COPPOLA - MERCK & CO., INC.
; STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000
; CITY: RAHWAY
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,007A
; FILING DATE: 13-FEB-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: COPPOLA, JOSEPH A
; REGISTRATION NUMBER: 38,413
; REFERENCE/DOCKET NUMBER: 19644
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732-594-6734
; TELEFAX: 732-594-4720
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 47 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

```

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-800-007A-7

Query Match          36.9%; Score 16.6; DB 3; Length 47;
Best Local Similarity 71.0%; Pred. No. 2.4e+02;
Matches 22; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 accgcagcgcaacttaaacacttggc 33
Db 44 ACCGCATGTCATCATCAACACCACTCATGAC 14

RESULT 11
US-08-171-389-382
; Sequence 382, Application US/08171389
; Patent No. 5578444
; GENERAL INFORMATION:
; APPLICANT: Edwards, Cynthia A.
; APPLICANT: Cantor, Charles R.
; APPLICANT: Andrews, Beth M.
; APPLICANT: Turin, Lisa M.
; APPLICANT: Fry, Kirk E.
; TITLE OF INVENTION: Sequence-Directed DNA Binding
; TITLE OF INVENTION: Molecules, Compositions and Methods
; NUMBER OF SEQUENCES: 641
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/171,389
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,936
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/723,618
; FILING DATE: 27-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,070
; FILING DATE: 22-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0175/G19P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 382:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Human 70 kDa heat shock protein gene
; US-08-171-389-382
```





```
; Patent No. 6010849
; GENERAL INFORMATION:
; APPLICANT: Edwards, Cynthia A.
; APPLICANT: Cantor, Charles R.
; APPLICANT: Andrews, Beth M.
; APPLICANT: Turin, Lisa M.
; APPLICANT: Fry, Kirk E.
; TITLE OF INVENTION: Sequence-Directed DNA Binding
; TITLE OF INVENTION: Molecules, Compositions and Methods
; NUMBER OF SEQUENCES: 664
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,080A
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/171,389
; FILING DATE: 20-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,936
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/723,618
; FILING DATE: 27-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,070
; FILING DATE: 22-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Brady, John F.
; REGISTRATION NUMBER: 39,118
; REFERENCE/DOCKET NUMBER: 4600-0175.20/G19P3D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 324-0880
; TELEFAX: (650) 324-0960
; INFORMATION FOR SEQ ID NO: 382:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Human 70 kDa heat shock protein gene
; US-08-482-080A-382

Query Match          36.4%; Score 16.4; DB 5; Length 50;
Best Local Similarity 67.6%; Pred. No. 2.9e+02;
Matches 23; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```
Oy 1 ggaagcattggcaacttaaaacatttgca 34
    ||| ||| ||| ||| ||| ||| |||
Db 4 ggtctccgtgacgacttataaaacccagggca 37
```

```
RESULT 15
PCT-US93-12388-382
; Sequence 382, Application PC/TUS9312388
; GENERAL INFORMATION:
```

```
; APPLICANT:
; TITLE OF INVENTION: Sequence-Directed DNA Binding
; TITLE OF INVENTION: Molecules, Compositions and Methods
; NUMBER OF SEQUENCES: 641
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/12388
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,936
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0175.41/G19PCT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 382:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Human 70 kDa heat shock protein gene
; PCT-US93-12388-382
```

```
Query Match          36.4%; Score 16.4; DB 6; Length 50;
Best Local Similarity 67.6%; Pred. No. 2.9e+02;
Matches 23; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```
Oy 1 ggaagcattggcaacttaaaacatttgca 34
    ||| ||| ||| ||| ||| ||| |||
Db 4 ggtctccgtgacgacttataaaacccagggca 37
```

```
Search completed: June 4, 2000, 16:09:46
Job time: 28065 sec
```

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:05:40 ; Search time 1236.38 Seconds

(Without alignments)  
-35.406 Million cell updates/sec

Title: US-09-164-714-16

Perfect score: 45  
Sequence: 1 atcaagcttagtgatgtga.....atgaagaagccaatgagcgc 45

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 356616

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_om:\*  
5: gb\_om:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl1:\*  
10: gb\_pl2:\*  
11: gb\_pl3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vl:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_om:\*  
23: em\_om:\*  
24: em\_om:\*  
25: em\_om:\*  
26: em\_om:\*  
27: em\_om:\*  
28: em\_om:\*  
29: em\_om:\*  
30: em\_om:\*  
31: em\_om:\*  
32: gb\_hg1:\*  
33: gb\_hg2:\*  
34: gb\_hg3:\*  
35: gb\_hg4:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pl4:\*  
41: gb\_hg5:\*  
42: gb\_hg6:\*  
43: gb\_hg7:\*  
44: gb\_hg8:\*

45: gb\_hg9:\*  
46: em\_hg1:\*  
47: em\_hg2:\*  
48: em\_hg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pl5:\*  
52: gb\_hg8:\*  
53: gb\_hg9:\*  
54: gb\_hg10:\*  
55: gb\_hg11:\*  
56: gb\_hg12:\*  
57: gb\_hg13:\*  
58: gb\_hg14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	28.8	64.0	48	5	AR024453	AR024453 Sequence
2	25.6	56.9	45	5	A77178	A77178 Sequence 6
3	25.6	56.9	45	5	A77180	A77180 Sequence 6
4	25.6	56.9	45	5	I20213	I20213 Sequence 6
5	25.6	56.9	45	5	I20215	I20215 Sequence 8
6	25.4	56.4	31	5	A02249	A02249 Oligonucleo
7	25.4	56.4	32	5	AR014147	AR014147 Sequence
8	25.4	56.4	32	5	AR014350	AR014350 Sequence
9	25.4	56.4	42	5	A77174	A77174 Sequence 2
10	25.4	56.4	42	5	I20209	I20209 Sequence 2
11	25.4	56.4	54	5	A77184	A77184 Sequence 12
12	25.4	56.4	54	5	I20219	I20219 Sequence 12
13	24.4	54.2	32	5	AR051772	AR051772 Sequence
14	24.4	49.8	31	5	AR051765	AR051765 Sequence
15	22.4	49.8	39	5	AR005129	AR005129 Sequence
16	22.4	49.8	39	5	AR058256	AR058256 Sequence
17	22.4	49.8	50	5	A68390	A68390 Sequence 7
18	22.4	49.8	50	5	A68391	A68391 Sequence 8
19	22.4	49.8	61	10	S71567	S71567 catechol O-
20	22.4	49.8	63	5	AR038968	AR038968 Sequence
21	22.2	49.3	37	5	A77195	A77195 Sequence 23
22	22.2	49.3	37	5	I20230	I20230 Sequence 23
23	22.2	48.9	38	5	A79123	A79123 Sequence 7
24	22.2	48.9	38	5	A79124	A79124 Sequence 8
25	22.2	48.9	38	5	A92205	A92205 Sequence 7
26	22.2	48.9	38	5	A92206	A92206 Sequence 8
27	21.8	48.4	37	5	A77194	A77194 Sequence 22
28	21.8	48.4	37	5	I20229	I20229 Sequence 22
29	21.4	47.6	31	5	A02248	A02248 Oligonucleo
30	21.4	47.6	32	5	AR014146	AR014146 Sequence
31	21.4	47.6	32	5	AR014349	AR014349 Sequence
32	21.4	47.6	32	5	AR051764	AR051764 Sequence
33	21.4	47.6	48	5	I23381	I23381 Sequence 24
34	21.4	47.6	48	5	I23383	I23383 Sequence 24
35	20.8	46.2	48	5	A69104	A69104 Sequence 22
36	20.4	45.3	42	5	I23963	I23963 Sequence 12
37	20.4	45.3	42	5	I23964	I23964 Sequence 13
38	20.4	45.3	60	5	AR016577	AR016577 Sequence
39	20.4	45.3	63	5	I23344	I23344 Sequence 15
40	20.4	45.3	63	5	A49833	A49833 Sequence 6
41	20.2	44.9	48	5	A92221	A92221 Sequence 23
42	20.2	44.4	23	5	A69105	A69105 Sequence 23
43	20.2	44.4	27	5	AR026052	AR026052 Sequence
44	19.8	44.0	30	5	A37681	A37681 Sequence 52
45	19.8	44.0	30	5	AR036846	AR036846 Sequence

## ALIGNMENTS

[illegible]

	Query Match	Score	DB	Length
Best Local Similarity	82.5%	Pred. No. 1.8,		
Matches	33;	Conservative	0;	Mismatches 7; Indels 0;
Qy	4	aagcttagtagatggtgtagtgaataaagccaatgagc	43	
Db	3	aagcttagtagatggtgtagtgaataaagccaatgagc	42	

[illegible]

Query Match	56.9%	Score 25.6;	DB 5;	Length 45;
Best Local Similarity	87.5%	Pred. No. 25;		
Matches	28;	Conservative	0;	Mismatches 4;
				Indels 0;
				Gaps
Qy	4	aagcttagtgbatggtgacgtgataaagcc	35	
Db	3	aagcttagtgcgtgcgtgcgtgcgtgcgtgc	34	
RESULT	3			
A77180				
LOCUS	A77180	45 bp	DNA	PAT
DEFINITION	Sequence 8 from Patent EP0614989.			19-OCT-1999
ACCESSION	A77180			
VERSION	A77180.1	GI:6088887		
KEYWORDS				
SOURCE	unidentified.			
ORGANISM	unidentified			
	unclassified.			

REFERENCE	1 (bases 1 to 45)
AUTHORS	Pleuecthun,A.D. and Kreibber,C.
TITLE	A METHOD FOR IN VIVO SELECTION OF LIGAND-BINDING PROTEINS
JOURNAL	Patent: EP 0614989-A 14-SEP-1994;
MORPHOSYS	PROTEINOPTIMERUNG (DE)
FEATURES	Location/Qualifiers
SOURCE	1..45
BASE COUNT	g a 7 c 13 g 16 t
ORIGIN	/organism="unidentified" /db_xref="taxon:32644"
Query Match	56.9% Score 25.6; DB 5; Length 45;
Best Local Similarity	87.5%; Pred. No. 25;
Matches	28; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Oy	4 aaagcttagatgatgtgatgtgaataaacgc 35       
Db	3 AAGCTTAGTGATGGATGCGTATGATGATTTC 34 

	RESULT	4	PAT	07-OCT-1996
	LOCUS	I20213		
	DEFINITION	I20213	45 bp	DNA
	ACCESSION	Sequence 6 from patent US 5514548.		
	VERSION	I20213		
	KEYWORDS	I20213.1	GI:1600568	
	SOURCE	.		
	ORGANISM	Unknown.		
	REFERENCE	Unclassified.		
	AUTHORS	1 (bases 1 to 45)		
	TITLE	Kreber,K., Moroney,S., Pluckthu,n,A. and Schneider,C.		
	JOURNAL	Method for in vivo selection of ligand-binding proteins		
	FEATURES	Patent: US 5514548 A 6 07-MAY-1996;		
	source	Location/Qualifiers		
		l..45		
BASE COUNT		/organism= "unknown"		
ORIGIN	g	a	c	t
	9	a	7 c	13 g
				16 t

Query Match	56.9%	Score 25.6;	DB 5,	Length 45;
Best Local Similarity	87.5%	Pred. NO.25;		
Matches	28;	Conservative	0;	Mismatches
			4;	Indels
				Gaps
				0.
0y	4	aagctagatgagtgatgagtgataaaagc	35	
Db	3	AAgCTTAGTGATGATGATGATGATGATTTC	34	
RESULT	5			
LOCUS	I20215			
DEFINITION	I20215	45 bp	DNA	
ACCESSION	I20215	Sequence 8 from patent US 5514548.	PAT	07-OCT-1996
VERSION	I20215.1	GI:1600570		
KEYWORDS	.			
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	1 (bases 1 to 45)			
AUTHORS	Krebbber, K., Moroney, S., Pluckthun, A. and Schneider, C.			
TITLE	Method for in vivo selection of ligand-binding proteins			
JOURNAL	Patent: US 5514548-A 8 07-MAY-1996;			
FEATURES	Location/Qualifiers			
source	1..45			
	/organism="unknown"			
BASE COUNT	9 a	7 c	13 g	16 t
ORIGIN				

Query Match 56.9%; Score 25.6; DB 5; Length 45;  
Best Local Similarity 87.5%; Pred. No. 25;  
Matches 28; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 aagcttagtgatggtgatgtaaaagcc 35  
|||||  
Db 3 AACCTTAGTGATGCTGATGATGATTC 34

RESULT 6  
LOCUS A02249 31 bp DNA PAT 26-APR-1996  
DEFINITION Oligonucleotide sequence (adaptor 9) from patent EP0282042.  
ACCESSION A02249  
VERSION A02249.1 GI:490303  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 31)  
AUTHORS Doebl, H., Eggmann, B., Gentz, R., Hochuli, E. and Stueber, D.  
TITLE Fusion proteins and their purification  
JOURNAL Patent: EP 0282042-A 23 14-SEP-1988;  
F. HOFMANN-LA ROCHE AG  
FEATURES Location/Qualifiers  
source 1..31  
/organism="synthetic construct"  
/db\_xref="taxon:32630"

BASE COUNT 7 a 2 c 12 g 10 t  
ORIGIN

Query Match 56.4%; Score 25.4; DB 5; Length 31;  
Best Local Similarity 96.3%; Pred. No. 29;  
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 agcttagtgatggtgatgtaaa 31  
|||||  
Db 1 AGCTTAGTGATGCTGATGATGAGA 27

RESULT 7  
LOCUS AR014147 32 bp DNA PAT 05-DEC-1998  
DEFINITION Sequence 113 from patent US 5773249.  
ACCESSION AR014147  
VERSION AR014147.1 GI:3971601  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 32)  
AUTHORS Cappello, J. and Ferrari, F.A.  
TITLE High molecular weight collagen-like protein polymers  
JOURNAL Patent: US 5773249-A 113 30-JUN-1998;  
FEATURES Location/Qualifiers  
source 1..32  
/organism="unknown"

BASE COUNT 9 a 2 c 12 g 9 t  
ORIGIN

Query Match 56.4%; Score 25.4; DB 5; Length 32;  
Best Local Similarity 96.3%; Pred. No. 29;  
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 agcttagtgatggtgatgtaaa 31  
|||||  
Db 1 AGCTTAGTGATGCTGATGATGAGA 27

RESULT 8  
LOCUS AR014350

LOCUS AR014350 32 bp DNA PAT 05-DEC-1998  
DEFINITION Sequence 25 from patent US 5773577.  
ACCESSION AR014350  
VERSION AR014350.1 GI:3971804  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 32)  
AUTHORS Cappello, J.  
TITLE Products comprising substrates capable of enzymatic cross-linking  
JOURNAL Patent: US 5773577-A 25 30-JUN-1998;  
FEATURES Location/Qualifiers  
source 1..32  
/organism="unknown"

BASE COUNT 9 a 2 c 12 g 9 t  
ORIGIN

Query Match 56.4%; Score 25.4; DB 5; Length 32;  
Best Local Similarity 96.3%; Pred. No. 29;  
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 agcttagtgatggtgatgtaaa 31  
|||||  
Db 1 AGCTTAGTGATGCTGATGATGAGA 27

RESULT 9  
LOCUS A77174 42 bp DNA PAT 19-OCT-1999  
DEFINITION Sequence 2 from Patent EP0614989.  
ACCESSION A77174  
VERSION A77174.1 GI:6088881  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 42)  
AUTHORS Plueckthun, A.D. and Krebber, C.  
TITLE A METHOD FOR IN VIVO SELECTION OF LIGAND-BINDING PROTEINS  
JOURNAL Patent: EP 0614989-A 14-SEP-1994;  
MORPHOSYS PROTEINOPTIMIERUNG (DB)  
FEATURES Location/Qualifiers  
source 1..42  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 11 a 7 c 15 g 9 t  
ORIGIN

Query Match 55.6%; Score 25; DB 5; Length 42;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 aagcttagtgatggtgatg 28  
|||||  
Db 3 AACCTTAGTGATGCTGATGATG 27

RESULT 10  
LOCUS I20209 42 bp DNA PAT 07-OCT-1996  
DEFINITION Sequence 2 from patent US 5514548.  
ACCESSION I20209  
VERSION I20209.1 GI:1600564  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 42)  
AUTHORS Krebber, K., Moroney, S., Plueckthun, A. and Schneider, C.  
TITLE Method for in vivo selection of ligand-binding proteins

JOURNAL Patent: US 5514548-A 2 07-MAY-1996;  
FEATURES Location/Qualifiers  
source 1..42  
/organism="unknown"  
BASE COUNT 11 a 7 c 15 g 9 t  
ORIGIN

Query Match 55.6%; Score 25; DB 5; Length 42;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 aagcttagtgatggtgatggtgatg 28  
|||||  
Db 3 AAGCTTAGTGATGCGTGTGATG 27

RESULT 11  
A77184 54 bp DNA PAT 19-OCT-1999  
LOCUS  
DEFINITION Sequence 12 from Patent EP0614989.  
ACCESSION A77184  
VERSION A77184.1 GI:6088891  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Pluckthun,A.D. and Krebber,C.  
TITLE A METHOD FOR IN VIVO SELECTION OF LIGAND-BINDING PROTEINS  
JOURNAL Patent: EP 0614989-A 14-SEP-1994;  
MORPHOSYS PROTEINOPTIMERUNG (DE)  
FEATURES Location/Qualifiers  
source 1..54  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 14 a 6 c 21 g 13 t  
ORIGIN

Query Match 55.6%; Score 25; DB 5; Length 54;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 aagcttagtgatggtgatggtgatg 28  
|||||  
Db 3 AAGCTTAGTGATGCGTGTGATG 27

RESULT 12  
I20219 54 bp DNA PAT 07-OCT-1996  
LOCUS  
DEFINITION Sequence 12 from patent US 5514548.  
ACCESSION I20219  
VERSION I20219.1 GI:1600574  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Krebber,K., Moroney,S., Pluckthun,A. and Schneider,C.  
TITLE Method for in vivo selection of ligand-binding proteins  
JOURNAL Patent: US 5514548-A 12 07-MAY-1996;  
FEATURES Location/Qualifiers  
source 1..54  
/organism="unknown"  
BASE COUNT 14 a 6 c 21 g 13 t  
ORIGIN

Query Match 55.6%; Score 25; DB 5; Length 54;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 aagcttagtgatggtgatggtgatg 28  
|||||  
Db 3 AAGCTTAGTGATGCGTGTGATG 27

RESULT 13  
AR051772 32 bp DNA PAT 29-SEP-1999  
LOCUS  
DEFINITION Sequence 99 from patent US 5830713.  
ACCESSION AR051772  
VERSION AR051772.1 GI:5975136  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 32)  
AUTHORS Ferrari,F.A., Cappello,J., Crissman,J.W. and Dorman,M.A.  
TITLE Methods for preparing synthetic repetitive DNA  
JOURNAL Patent: US 5830713-A 99 03-NOV-1998;  
FEATURES Location/Qualifiers  
source 1..32  
/organism="unknown"  
BASE COUNT 8 a 2 c 12 g 10 t  
ORIGIN

Query Match 54.2%; Score 24.4; DB 5; Length 32;  
Best Local Similarity 96.2%; Pred. No. 67;  
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 6 gcttagtgatggtgatggtgatgaa 31  
|||||  
Db 2 GCTTAGTGATGCGTGTGATGAGA 27

RESULT 14  
AR051765 31 bp DNA PAT 29-SEP-1999  
LOCUS  
DEFINITION Sequence 83 from patent US 5830713.  
ACCESSION AR051765  
VERSION AR051765.1 GI:5975129  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 31)  
AUTHORS Ferrari,F.A., Cappello,J., Crissman,J.W. and Dorman,M.A.  
TITLE Methods for preparing synthetic repetitive DNA  
JOURNAL Patent: US 5830713-A 83 03-NOV-1998;  
FEATURES Location/Qualifiers  
source 1..31  
/organism="unknown"  
BASE COUNT 9 a 1 c 12 g 9 t  
ORIGIN

Query Match 49.8%; Score 22.4; DB 5; Length 31;  
Best Local Similarity 95.8%; Pred. No. 3.5e+02;  
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 8 tttagtgatggtgatggtgatgaa 31  
|||||  
Db 3 TTAGTGATGCGTGTGATGAGA 26

RESULT 15  
AR005129 39 bp DNA PAT 04-DEC-1998  
LOCUS  
DEFINITION Sequence 34 from patent US 5747452.  
ACCESSION AR005129  
VERSION AR005129.1 GI:3966008  
KEYWORDS

Unknown.  
Unknown.

1 (bases 1 to 39)

### Method of modulating tumor

III peptides  
Retent.: no 5737450 x 34 of www 1000

Patent: US 5747452-A 34 05-MAY-1998;  
Location/Qualifiers  
1 30

1. .39  
/organism="unknown"

13 a 15 c 3 g 8 t

ch 49.8%; Score 22.4; DB 5; Length 39;  
1 Similarity 95.8%; Pred. No. 3.5e+02;  
23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

[illegible]

leted: June 4, 2000, 16:05:41  
7893 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 13:53:25 ; Search time 4521.53 Seconds  
(without alignments)  
40.339 Million cell updates/sec

Title: US-09-164-714-16

Perfect score: 45

Sequence: 1 atcaagcttgatgtagtga.....atgaagccaatgagcgc 45

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 202661650 residues

Total number of hits satisfying chosen parameters: 119968

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gp\_est1:\*  
21: gp\_est2:\*  
22: gp\_est3:\*  
23: gp\_est4:\*  
24: gp\_est5:\*  
25: gp\_est6:\*  
26: gp\_est7:\*  
27: gp\_est8:\*  
28: gp\_est9:\*  
29: gp\_est10:\*  
30: gp\_est11:\*  
31: gp\_est12:\*  
32: gp\_est13:\*  
33: gp\_est14:\*  
34: gp\_est15:\*  
35: gp\_est16:\*  
36: gp\_est17:\*  
37: gp\_est18:\*  
38: gp\_est19:\*  
39: gp\_est20:\*  
40: gp\_est21:\*  
41: gp\_est22:\*  
42: gp\_est23:\*  
43: gp\_est24:\*  
44: gp\_est25:\*

45: gp\_est26:\*  
46: gp\_est27:\*  
47: gp\_est28:\*  
48: gp\_est29:\*  
49: gp\_est30:\*  
50: gp\_est31:\*  
51: gp\_est32:\*  
52: gp\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gp\_est33:\*  
60: gp\_est34:\*  
61: gp\_est35:\*  
62: gp\_est36:\*  
63: gp\_est37:\*  
64: gp\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gp\_est39:\*  
70: gp\_est40:\*  
71: gp\_est41:\*  
72: gp\_est42:\*  
73: gp\_est43:\*  
74: gp\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gp\_est45:\*  
80: gp\_est46:\*  
81: gp\_est47:\*  
82: gp\_est48:\*  
83: gp\_est49:\*  
84: gp\_est50:\*  
85: gp\_est51:\*  
86: em\_est35:\*  
87: em\_est36:\*  
88: em\_est37:\*  
89: em\_est38:\*  
90: gp\_est52:\*  
91: gp\_est53:\*  
92: gp\_est54:\*  
93: gp\_est55:\*  
94: gp\_est56:\*  
95: em\_est39:\*  
96: em\_est40:\*  
97: em\_est41:\*  
98: em\_est42:\*  
99: em\_est43:\*  
100: em\_est44:\*  
101: em\_est45:\*  
102: gp\_est57:\*  
103: gp\_est58:\*  
104: em\_est46:\*  
105: gp\_est59:\*  
106: gp\_est60:\*  
107: gp\_est61:\*  
108: gp\_est62:\*  
109: gp\_est63:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result

Query

SUMMARIES

No.	Score	Match	Length	DB	ID	Description
C 1	20	44.4	100	94	A0026191	A0026191 1(3)03847
C 2	18.8	40.4	76	44	A1313567	A1313567 EST021 Mo
C 3	18.2	40.4	90	51	A0025298	A0025298 EP(3)3123
C 4	18.2	40.4	90	51	A1748511	A1748511 sb5fa04.y
C 5	18.2	40.4	90	69	AM119559	AM119559 sdd8a04.y
C 6	18.2	40.4	91	45	A1366782	A1366782 qg4ld10.x
C 7	18.2	40.4	100	39	AA906127	AA906127 o39c04.s
C 8	17.8	39.6	83	28	AA108264	AA108264 EST0007 r
C 9	17.6	39.1	73	50	A1708585	A1708585 as98g06.x
C 10	17.6	39.1	84	27	AA038117	AA038117 m181g04.r
C 11	17.4	38.7	58	36	AA663265	AA663265 ab80f10.s
C 12	17.4	38.7	87	39	AA861714	AA861714 ak30h09.s
C 13	17.4	38.7	88	47	A1522508	A1522508 fb20d05.x
C 14	17.2	38.2	89	23	L46917	L46917 SCMRAP004 G
C 15	17.2	38.2	100	63	A1963567	A1963567 wr64h11.x
C 16	17	37.8	84	83	AF039763	AF039763 Homo sapi
C 17	17	37.8	100	35	AA537389	AA537389 vj98a07.r
C 18	16.8	37.3	65	41	AU008149	AU008149 AU008149
C 19	16.8	37.3	88	39	AA885639	AA885639 o332b08.s
C 20	16.8	37.3	94	46	A1460411	A1460411 sa81c02.y
C 21	16.8	37.3	100	39	AA902339	AA902339 OK92c10.s
C 22	16.8	37.3	100	62	A1921625	A1921625 wo27h05.x
C 23	16.8	37.3	100	62	A1921625	A1921625 wo27h05.x
C 24	16.6	36.9	40	45	AU237265	AU237265 AJ237265
C 25	16.6	36.9	59	27	AA026247	AA026247 z399a04.s
C 26	16.6	36.9	63	82	GA0246098	GA0246098 Gallus ga
C 27	16.6	36.9	71	44	A1310271	A1310271 tba47h04.x
C 28	16.6	36.9	82	68	HSMD002994	HSMD002994 Homo sapi
C 29	16.6	36.9	86	35	AA574367	AA574367 nt46b08.s
C 30	16.6	36.9	86	82	CNS00W22	AL0944220 Arabidops
C 31	16.6	36.9	88	68	HSMD002995	AL038519 Homo sapi
C 32	16.4	36.4	61	33	AA417038	AA417038 zui3a02.s
C 33	16.4	36.4	85	26	W99173	W99173 ME91e04.r1
C 34	16.4	36.4	85	33	AA443226	AA443226 aal1a08.r
C 35	16.4	36.4	90	50	AU060644	AU060644 AU060644
C 36	16.4	36.4	91	21	T96074	T96074 ye47a12..s1
C 37	16.4	36.4	96	84	B34289	B34289 HS-1024-A2-
C 38	16.2	36.0	46	37	AA714376	AA714376 nw20b06.s
C 39	16.2	36.0	65	81	AA423329	AA423329 sh66c03.y
C 40	16.2	36.0	67	63	A1037391	A1037391 ub52b11.r
C 41	16.2	36.0	67	63	A1988376	A1988376 s01b03.y
C 42	16.2	36.0	70	34	AA467126	AA467126 vd98d01.r
C 43	16.2	36.0	71	45	A1348869	A1348869 t05b07.x
C 44	16.2	36.0	73	33	AA445034	AA445034 v957g06.r
C 45	16.2	36.0	73	34	AA518487	AA518487 v101b09.r

## ALIGNMENTS

RESULT 1  
LOCUS A0026191/c 100 bp DNA 30-JUN-1998  
DEFINITION I(3)03847 Drosophila melanogaster P lethal line Drosophila  
melanogaster genomic Sequence recovered from Both 5' and 3' ends of  
P element, genomic survey sequence.

ACCESSION A0026191  
VERSION A0026191.1 GI:3266476  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 100)  
A0026191 Spradling,A.C., Stern,D., Beaton,A., Rehm,E.J., Laverly,T.,  
Mozden,N., Mistra,S. and Rubin,G.M.  
The BDP gene disruption project: Single P element insertions  
mutating 30% of Drosophila autosomal genes  
JOURNAL Unpublished (1998)  
COMMENT Contact: Gerald Rubin

Berkeley Drosophila Genome Project  
University of California, Berkeley  
LSA Building, Berkeley, CA 94720-3200, USA  
Fax: 5106439947  
Email: gerry@fruitfly.berkeley.edu  
Sequence recovery method was inverse PCR.  
Sequence orientation is forward strand relative to 5' end of P  
element  
The P element insertion position is base 052 in the 100 bases. This  
insertion position refers to the first base of the 8 base target  
recognition sequence.

FEATURES  
source location/Qualifiers  
1..100  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="Drosophila melanogaster P lethal line"  
/note="Inverse PCR was performed on Drosophila  
melanogaster strains each of which contains a single P  
transposable element insertion that is thought to cause  
either lethality or sterility. The resultant fragment for  
each strain was directly sequenced to determine the  
genomic sequence at the site of insertion. Details of the  
protocols used can be found at  
http://fruitfly.berkeley.edu/P-distrupt/Inverse\_Pcr.html."

BASE COUNT 20 a 35 c 21 g 24 t  
ORIGIN

Query Match 44.4% Score 20; DB 94; Length 100;  
Best Local Similarity 65.9%; Pred No. 1.6e+03;  
Matches 29; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 2 tcaagcttagtgatgtagtgaagaaagccaatgagc 45  
||||| ||||| || || ||||| || |||||  
DB 44 tcaagaaatgtagtgaagaaagccaatgagc 1

RESULT 2  
LOCUS A1313567/c 68 bp mRNA EST 15-DEC-1998  
DEFINITION EST021 Mouse ES cell lambda zap Express Library Mus musculus cDNA  
clone HBMG015r 3', mRNA sequence.  
ACCESSION A1313567  
VERSION A1313567.1 GI:4027734  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 68)  
A1313567 Kim,S.J., Shin,J.H., Kim,J.H., Kim,S.H., Chae,J.H., Park,E.J.,  
Hong,S.H., Park,S.D., Jeong,S.J. and Kim,C.G.  
Isolation of Developmentally Regulated Novel Genes Based on  
Sequence Identity and Gene Expression Pattern  
Unpublished (1998)

JOURNAL On Aug 26, 1998 this sequence version replaced gi:3475005.  
COMMENT Contact: Chul Geun Kim

Molecular Genetics  
University of Hanyang, Department of Life Science  
Haengdang dong 17, Sungdong gu, Seoul 133-791, Korea (ROK)  
Tel: 82-2-2290-0957  
Fax: 82-2-2296-5996  
Email: cgm@emall.hanyang.ac.kr  
Randomly chosen cDNA clone from mouse ES cell cDNA library with  
homology to XPRC/ERC-3 (DNA repair gene)  
Seq primer: 73 Reverse  
High quality sequence stop: 68.  
Location/Qualifiers

```

source
1. 68
/organism="Mus musculus"
/strain="129/sv"
/db_xref="taxon:10090"
/clone_lib="HBMG015R"
/clone_id="Mouse ES cell Lambda Zap Express Library"
/sex="male"
/tissue_type="blastocyst inner cell mass"
/cell_type="embryonic stem cell"
/dev_stage="blastocyst"
/notes="Vector: Uni-Zap XR Phage and generate
pbuascript-SK phagemid vector by in vivo excision."
BASE COUNT      19 a      18 c      13 g      18 t
ORIGIN

Query Match      41.8%; Score 18.8; DB 44; Length 68;
Best Local Similarity 76.7%; Pred. No. 3.9e+03;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      1 atcaagcttagtgatggtgatgataa 30
      1 ||| ||| ||| ||| ||| ||| |||
      50 ACCAAGCTTGAGATGAAGATGGTGTGAA 21

RESULT 3
A0025298      76 bp      DNA      GSS      14-OCT-1998
LOCUS      EP(3)123-Sprime Drosophila melanogaster EP line Drosophila
DEFINITION      melanogaster genomic Sequence recovered from 5' end of P element,
                genomic survey sequence.
ACCESSION      A0025298
VERSION      A0025298.1 GI:3265650
KEYWORDS      GSS.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
                Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
                Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
                Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 76)
Rehm,E.J. and Rubin,G.M.
The BDGP gene disruption project: single EP element insertions
unpublished (1998)
Contact: Gerald Rubin
Berkeley Drosophila Genome Project
University of California, Berkeley
LSA Building, Berkeley, CA 94720-3200, USA
Fax: 5106439947
Email: gerry@fruitfly.berkeley.edu

Sequence recovery method was inverse PCR.

element orientation is forward strand relative to 5' end of P
element

The P element insertion position is base 69 in the 76 bases. This
insertion position refers to the first base of the 8 base target
recognition sequence.

Class: transposon-tagged.
Location/Qualifiers
1. 76
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="Drosophila melanogaster EP line"
/notes="Inverse PCR was performed on Drosophila
melanogaster strains each of which contains a single EP
transposable element insertion. (The generation of these
insertion strains is described in Rorth P, Szabo K, Bailey
A, Laverly T, Rehm J, Rubin GM, Weigmann K, Millan M, Benes
V, Ansoorge W, Cohen SM. 1998. Systematic gain-of-function
genetics in Drosophila. Development 6:1049-1057.) The
resultant fragment for each strain was directly sequenced

```

```

to determine the genomic sequence at the site of
insertion. Details of the protocols used can be found at
http://fruitfly.berkeley.edu/P_disrupt/inverse_por.html."
BASE COUNT      18 a      14 c      19 g      25 t
ORIGIN

Query Match      40.4%; Score 18.2; DB 94; Length 76;
Best Local Similarity 74.2%; Pred. No. 6.4e+03;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY      8 ttatgatgtagtgatgataaagccaa 38
      1 ||| ||| ||| ||| ||| ||| |||
      45 TTTCGTCTTGATGTGAATGAAGCCAAA 75

RESULT 4
A1748511/c      90 bp      mRNA      EST      13-DEC-1999
LOCUS      sb54s04.y1 Gm-cl016 glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION      Gm-cl016-295 5', mRNA sequence.
ACCESSION      A1748511
VERSION      A1748511.1 GI:5126775
KEYWORDS      EST.
SOURCE      soybean.
ORGANISM      Glycine max
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
                eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoidae;
                Glycine.
1 (bases 1 to 90)
Shoemaker,R., Kein,P., Vodkin,L., Erpelting,J., Coryell,V.,
Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
Bowers,Y., Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N.,
Schurr,R., Ritter,E., Kohn,S., Shi,T., Jackson,Y., Cardenas,M.,
McCann,R., Waterson,R. and Wilson,R.
Public Soybean EST Project
unpublished (1999)
On Jun 5, 1998 this sequence version replaced gi:3188332.
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com.

FEATURES
source
1. 90
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl016-295"
/clone_id="Gm-cl016"
/tissue_type="immature flowers of field grown plants"
/lab_host="XL10-Gold"
/notes="Vector: pBluescript II XR. Site_1: EcoRI, Site_2:
XhoI; This cDNA library was constructed from mRNA isolated
from immature flowers of field grown plants. The cDNA
library was prepared using the Stratagene pBluescript II
XR library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dT) sequence with a XhoI restriction site. EcoRI adapters
were ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into XL10-Gold host cells. This library was
constructed by Dr. Randy Shoemaker and Dr. John

```

BASE COUNT	24 a	21 c	11 g	34 t
ORIGIN	Erpelding.."			
Query Match	Best Local Similarity	66.7%;	Score 18.2;	DB 51; Length 90;
Matches	26; Conservative	0;	Mismatches 13;	Indels 0; Gaps 0;
Oy	4	aagcttagtgcgtgcatgcatgaagaacccaatgatg	42	
Dd	76	ACGCTTGTTGGTTATGCTTGCAGAAATATCCAAAACAG	38	
RESULT	5			
LOCUS	AM119559/c			
DEFINITION	AM119559	90 bp mRNA	EST	13-DEC-1999
ACCESSION	s448a04.y1	Gm-cl016 Glycine max CDNA clone	GENOME SYSTEMS CLONE ID:	
VERSION	Gm-cl016-2983	5', mRNA sequence.		
KEYWORDS	AM119559			
SOURCE	AM119559.1	GI:6094892		
ORGANISM	Esr.			
	soybean.			
	Glycine max			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.			
AUTHORS	1 (bases 1 to 90) Shoenmaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Holla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurr,R., Ritter,E., Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R. Public Soybean EST Project Unpublished (1999)			
TITLE	On Apr 30, 1999, this sequence version replaced gi:4727921.			
JOURNAL	Contact: Shoemaker R/Public Soybean EST project			
COMMENT	Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3322 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or Info@genomesystems.com web site: www.genomesystems.com Seq primer: -40RP from Glbcoc. Location/Qualifiers 1..90 /organism="Glycine max" /db_xref="taxon:3847" /clone="GENOME SYSTEMS CLONE ID: Gm-cl016-2983" /clone_1lb="Gm-cl016" /feature_type="immature flowers of field grown plants" /lab_host="X110-Gold" /note="Vector: pBluescript II XR; Site_1: EcoRI, Site_2: XhoI; This cDNA library was constructed from mRNA isolated from immature flowers of field grown plants. The cDNA library was prepared using the Strategene pBluescript II XR library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into X110-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelding."			

	ORIGIN	24 a	21 c	11 g	34 t
Qy	Query Match Best Local Similarity Matches	40.4%; 66.7%; 26;	Score 18.2; Pred. No. 6.5e+03; 0;	DB 69; Mismatched 13;	Length 90;  Indels 0; Gaps 0;
Dd	76 ACGCTTGTGTTGGTTAGTGTCAGAAATATCCAAAACAG 38	4 aagcttagtgcgtgcatgcatgaagaacccaatgag 42 1   26			
RESULT 6	LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS TITLE JOURNAL COMMENT	A1366782 91 bp mRNA EST 13-FEB-1999 gg41dd10.x1 Soares_NHMPU_S1 Homo sapiens CDNA clone IMAGE:1935091 3', mRNA sequence. A1366782 A1366782.1 GI:4136527 EST. human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheraia; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 91) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) On Jan 14, 1998 this sequence version replaced gi:1798622. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 725 Std Error: 0.00 Seq primer: -40UP from Glbpco High quality sequence stop: 71.			
FEATURES	source	1..91 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1935091" /clone_id="Soares_NHMPU_S1" /isue_type="Pooled human melanocyte, fetal heart, and pregnant uterus" /lab_host="DH10B" /note="Organ: mixed (see below); Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NBHM, pregnant uterus NBHFU, and fetal heart NBHI129W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."			
BASE COUNT	35 a 10 c 17 g 29 t				
ORIGIN					
Omy	Query Match Best Local Similarity Matches	40.4%; 66.7%; 26;	Score 18.2; Pred. No. 6.6e+03; 0;	DB 45; Mismatched 13;	Length 91;  Indels 0; Gaps 0;
Dd	51 CAGCATGTATGTGGAGAAGGTTAAGTACACCACATCA 89	3 caagcttagtgcgtgcatgcatgaagaacccaatga 41 1   26			
RESULT 7					

AA906127	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES
AA906127	100 bp	mRNA	EST	19-MAY-1998									
OJ93G04.S1	Scares_NFL_T_GBC_S1	Homo sapiens CDNA clone IMAGE:1505862 3' similar to SW:U0F4.CANFA Q95157 OLFATORY RECEPTOR-LIKE PROTEIN OLF4, ;, mRNA sequence.											
AA906127													
AA906127.1	GI:3041250	EST.											
		human.											
		Homo sapiens											
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.											
	1 (bases 1 to 100)												
	NCI-CCAP	http://www.ncbi.nlm.nih.gov/ncicgap.											
		National Cancer Institute, Cancer Genome Anatomy Project (CGAP),											
		Tumor Gene Index											
	Unpublished (1997)												
	On May 9, 1995	this sequence version replaced gi:802984.											
	Contact:	Robert Strausberg, Ph.D.											
	Tel: (301) 496-1550												
	Email:	Robert_Strausberg@nih.gov											
	This clone is available royalty-free through LIND, ; contact the												
	IMAGE Consortium (info@image.llnl.gov) for further information.												
	Trace considered overall poor quality												
	Insert Length: 833	Std Error: 0.00											
	Seq primer: -40m13 fwd. ET from Amersham												
	High quality sequence stop: 1.												
	Location/Qualifiers												
	1..100												
	/organism="Homo sapiens"												
	/db_xref="taxon:9606"												
	/clone="IMAGE:1505862"												
	/clone_11b="Scares_NFL_T_GBC_S1"												
	/lab_host="DH10B"												
	/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBL19W, testis NHT, and B-cell NCI-CGAP-GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1.M.A.G.E. clones 297480-302087, 682632-687239, 72608-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."												
	BASE COUNT	27 a	19 c	33 g	21 t								
	ORIGIN												
	Query Match	40.4%;	Score 18.2;	DB 39;	Length 100;								
	Best Local Similarity	74.2%;	Pred. No. 6.7+03;										
	Matches	23;	Conservative	0;	Mismatches	8;	Indels	0;	Gaps	0;			
OY	12	tgatgtgatgtgatgaagaagccaatgaag	42										
DB	1	TCAGGTCATCGAGAGAACCCCAACGAG	31										
RESULT	8												
LOCUS	AA108264/C	83 bp	mRNA	EST	05-SEP-1997								
DEFINITION	AA108264	AA108264	AA108264	AA108264	AA108264								
ACCESSION	AA108264	AA108264	AA108264	AA108264	AA108264								
VERSION	AA108264.1	GI:1659732											
KEYWORDS	EST.												
SOURCE	Norway rat.												
ORGANISM	Rattus norvegicus												
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.												
REFERENCE	1 (bases 1 to 83)					</							

TITLE	Rebillard, G., Vago, P., Pujol, R. and Hamel, C.P.		
JOURNAL	Identification of preferentially expressed cochlear genes by systematic sequencing of a rat cochlea cDNA library		
MEDLINE	Brain Res. Mol. Brain Res. 47 (1-2), 1-10 (1997)		
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1393651. Contact: Hamel, CP Laboratoire de Neurobiologie de l'Audition - INSERM U254 Institut National de la Sante et de la Recherche Medicale C.H.R. St-Charles, 34295 Montpellier cedex, France Tel: (33) 04 67 33 69 75 Fax: (33) 04 67 52 56 01 Email: biomol@met.fr		
FEATURES	PCR Primers FORWARD: pUC/M13 reverse primer BACKWARD: pUC/M13 forward primer Seq primer: pUC/M13 reverse primer High quality sequence stop: 83. Location/Qualifiers 1..83 /organism="Rattus norvegicus" /strain="Wistar" /db_xref="taxon:10116" /clone="PCO38" /clone_lib="rat lambda ZAPII library (C.P.Hamel)" /dev_stage="postnatal day 24" /lab_host="Escherichia coli" /note="Organ: cochlea; Vector: lambda ZAPII; Site_1: EcoRI; Site_2: XhoI"		
BASE COUNT	13 a	34 c	17 g
ORIGIN	18 t	1 others	
Query Match	39.6%	Score 17.8;	DB 28; Length 83;
Best Local Similarity	67.6%	Pred. No. 8.9e+03;	
Matches	25;	Conservative 0;	Mismatches 12; Indels 0; Gaps 0;
Db	45	CATGGTGTGGCGATGATGAGAAAGAGCGCTGAGGGGC	9
Oy	7	cttagtgatgltgatgtgatgaagaacaaatgagc	43
LOCUS	1		
DEFINITION	as98906.x1	Barstead aorta HPLB6	Homo sapiens CDNA clone
REGION (HUMAN);	3'	similar to gb:U03555	IG KAPPA CHAIN PRECURSOR V-I
ACCESSION	AI708585	EST	04-JUN-1999
VERSION	AI708585.1	GI:4998361	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 73)		
JOURNAL	Hillier, L., Allen, M., Bowles, L., Dubuque, T., Giesel, G., Jost, S.,		
COMMENT	Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,		
	Martin, J., Moore, B., Schellenberg, K., Stepien, M., Tan, F.,		
	Thaïsling, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.		
	WashU-NCI human EST Project		
	Unpublished (1997)		
	On Dec 20, 1995 this sequence version replaced gi:1133755.		
	Contact: Wilson RK		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@watson.wustl.edu		
	This clone is available royalty-free through LNL; contact the		
	IMAGE Consortium (info@image.lnl.gov) for further information.		
	Seq primer: -400P from Glpco		
	High quality sequence stop: 1.		









---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:09:46 ; Search time 244.64 Seconds  
(without alignments)  
23.910 Million cell updates/sec

Title: US-09-164-714-16

Perfect score: 45

Sequence: 1 atcaagcttgatgtgtga.....atgaagccaatgagcgc 45

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 375880

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/PCRUS.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/1na/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28.8	64.0	48	2	US-08-621-081A-15
2	25.6	56.9	45	1	US-08-197-770-6
3	25.6	56.9	45	1	US-08-197-770-8
4	25.4	56.4	32	2	US-08-642-255-113
5	25.4	56.4	32	2	US-08-397-633A-25
6	25.5	55.6	42	1	US-08-197-770-2
7	25.5	55.6	54	1	US-08-197-770-12
8	24.6	54.7	64	3	US-08-697-631A-12
9	24.6	54.7	65	3	US-08-697-631A-13
10	24.6	54.7	66	3	US-08-697-631A-11
11	24.4	54.2	32	3	US-08-697-631A-14
12	24.4	54.2	32	3	US-08-707-237A-99
13	24.4	54.2	32	5	US-08-642-246-21
14	24.4	54.2	32	6	PCR-US96-06229-21
15	22.4	49.8	31	2	US-08-707-237A-83
16	22.4	49.8	31	2	US-08-459-064B-34
17	22.4	49.8	39	3	US-08-460-421A-34
18	22.4	49.8	63	3	US-08-416-336-3
19	22.2	48.3	37	1	US-08-197-770-23
20	21.8	48.4	37	1	US-08-197-770-22
21	21.4	47.6	32	2	US-08-642-255-112
22	21.4	47.6	32	2	US-08-397-633A-24
23	21.4	47.6	32	2	US-08-707-237A-82
24	21.4	47.6	32	5	US-08-642-246-20
25	21.4	47.6	32	6	PCR-US96-06229-20
26	21.4	47.6	54	5	US-08-989-370-10
27	21.4	46.7	48	1	US-08-221-817-24

28	21	46.7	48	1	US-08-454-439-24	Sequence 24, Appl
29	21	46.7	48	6	PCR-US94-10487-24	Sequence 24, Appl
30	21	46.7	54	5	US-08-989-370-11	Sequence 11, Appl
31	20.6	45.8	54	5	US-08-989-370-7	Sequence 7, Appl
32	20.4	45.3	42	1	US-08-250-859-12	Sequence 12, Appl
33	20.4	45.3	42	1	US-08-250-859-13	Sequence 13, Appl
34	20.4	45.3	42	2	US-08-490-803-13	Sequence 12, Appl
35	20.4	45.3	42	2	US-08-490-803-13	Sequence 12, Appl
36	20.4	45.3	42	4	US-08-457-254-23	Sequence 23, Appl
37	20.4	45.3	42	4	US-08-457-254-24	Sequence 24, Appl
38	20.4	45.3	42	4	US-08-484-257-12	Sequence 12, Appl
39	20.4	45.3	42	4	US-08-484-257-13	Sequence 13, Appl
40	20.4	45.3	42	6	PCR-US94-08806-112	Sequence 12, Appl
41	20.4	45.3	42	6	PCR-US94-08806-113	Sequence 13, Appl
42	20.4	45.3	42	6	PCR-US95-01775-12	Sequence 12, Appl
43	20.4	45.3	42	6	PCR-US95-01775-13	Sequence 13, Appl
44	20.4	45.3	42	6	PCR-US95-16626-23	Sequence 23, Appl
45	20.4	45.3	42	6	PCR-US95-16626-24	Sequence 24, Appl

## ALIGNMENTS

RESULT 1

US-08-621-081A-15  
; Sequence 15, Application US/08621081A

; Patent No. 5795974

; GENERAL INFORMATION:

; APPLICANT: Cole, Barry C.

; APPLICANT: Knudson, Kevin L.

; APPLICANT: Sawitzke, Allen D.

; TITLE OF INVENTION: Mycoplasma Arthritis Superantigen

; NUMBER OF SEQUENCES: 32

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Thorpe, No. 5795974th & Western

; STREET: 9035 South 700 East, Suite 200

; CITY: Sandy

; STATE: Utah

; COUNTRY: USA

; ZIP: 84070

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage

; COMPUTER: IBM ThinkPad 340

; OPERATING SYSTEM: MS-DOS 6.22

; SOFTWARE: Word Perfect 6.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/621,081A

; FILING DATE: 424

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/165,038

; FILING DATE: December 10, 1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Alan J. Howarth

; REGISTRATION NUMBER: 36,553

; REFERENCE/DOCKET NUMBER: T676.CIP/U-1195

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (801)566-6633

; TELEFAX: (801)566-0750

; INFORMATION FOR SEQ ID NO: 15:

; LENGTH: 48

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-621-081A-15

Query Match 64.0%; Score 28.8; DB 2; Length 48;  
Best Local Similarity 82.5%; Pred. No. 0.02;  
Matches 33; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 aagcttagtgatggtgatgtaaaagccaatgagc 43  
|||||  
Db 3 AAGCTTAGTGATGGTGTGATGATGATCTTCAAAAACAC 42

## RESULT 2

US-08-197-770-6  
; Sequence 6, Application US/08197770  
; Patent No. 5514548  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: A Method for In Vivo Selection of  
; TITLE OF INVENTION: Ligand-Binding Proteins  
; NUMBER OF SEQUENCES: 23  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/197,770  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 93 10 2484.8  
; FILING DATE: 17-FEB-1993  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 45 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-197-770-6

Query Match 56.9%; Score 25.6; DB 1; Length 45;  
Best Local Similarity 87.5%; Pred. No. 0.28;  
Matches 28; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 aagcttagtgatggtgatgtaaaagcc 35  
|||||  
Db 3 AAGCTTAGTGATGGTGTGATGATGATTTCC 34

## RESULT 3

US-08-197-770-8  
; Sequence 8, Application US/08197770  
; Patent No. 5514548  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: A Method for In Vivo Selection of  
; TITLE OF INVENTION: Ligand-Binding Proteins  
; NUMBER OF SEQUENCES: 23  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/197,770  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 93 10 2484.8  
; FILING DATE: 17-FEB-1993  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 45 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single

; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-197-770-8

Query Match 56.9%; Score 25.6; DB 1; Length 45;  
Best Local Similarity 87.5%; Pred. No. 0.28;  
Matches 28; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 aagcttagtgatggtgatgtaaaagcc 35  
|||||  
Db 3 AAGCTTAGTGATGGTGTGATGATGATTTCC 34

## RESULT 4

US-08-642-255-113  
; Sequence 113, Application US/08642255  
; Patent No. 5773249  
; GENERAL INFORMATION:  
; APPLICANT: CAPPELLO, Joseph  
; TITLE OF INVENTION: FERRARI, Franco A.  
; TITLE OF INVENTION: High Molecular Weight Collagen-Like  
; NUMBER OF SEQUENCES: 135  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/642,255  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ROWLAND, Berttram I.  
; REGISTRATION NUMBER: 20, 015  
; REFERENCE/DOCKET NUMBER: A55556-3/BIR  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 494-8700  
; TELEFAX: (415) 494-8771  
; TELEX: 910 277299 FHT UR  
; INFORMATION FOR SEQ ID NO: 113:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 32 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "synthetic"  
US-08-642-255-113

Query Match 56.4%; Score 25.4; DB 2; Length 32;  
Best Local Similarity 96.3%; Pred. No. 0.31;  
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 aagcttagtgatggtgatgtaaa 31  
|||||  
Db 1 AGCTTAGTGATGGTGTGATGATGAGA 27

## RESULT 5

US-08-397-633A-25  
; Sequence 25, Application US/08397633A

```
; Patent No. 5773577
; GENERAL INFORMATION:
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATE SCAPABLE
; TITLE OF INVENTION: OF ENZYMA TIC CROSS-LINKING
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/397,633A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Berttram I
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROP-011-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-397-633A-25

Query Match          56.4%; Score 25.4; DB 2; Length 32;
Best Local Similarity 96.3%; Pred. No. 0.31;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 aagcttagtgatggtgatggtgataa 31
Db 1 aagcttagtgatggtgatggtgataa 27

RESULT 6
US-08-197-770-2
; Sequence 2, Application US/08197770
; Patent No. 5514548
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: A Method for In Vivo Selection of
; TITLE OF INVENTION: Ligand-Binding Proteins
; NUMBER OF SEQUENCES: 23
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/197,770
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 93 10 2484.8
; FILING DATE: 17-FEB-1993
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-197-770-2
```

```
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-197-770-2

Query Match          55.6%; Score 25; DB 1; Length 42;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 aagcttagtgatggtgatggtgatg 28
Db 3 aagcttagtgatggtgatggtgatg 27

RESULT 7
US-08-197-770-12
; Sequence 12, Application US/08197770
; Patent No. 5514548
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: A Method for In Vivo Selection of
; TITLE OF INVENTION: Ligand-Binding Proteins
; NUMBER OF SEQUENCES: 23
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/197,770
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 93 10 2484.8
; FILING DATE: 17-FEB-1993
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-197-770-12

Query Match          55.6%; Score 25; DB 1; Length 54;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 aagcttagtgatggtgatggtgatg 28
Db 3 aagcttagtgatggtgatggtgatg 27

RESULT 8
US-08-697-631A-12
; Sequence 12, Application US/08697631A
; Patent No. 5879673
; GENERAL INFORMATION:
; APPLICANT: Thomas, Griffith R
; TITLE OF INVENTION: No. 5879673el Administration of Thrombopoietin
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
```

```

; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/697,631A
; FILING DATE: 28-Aug-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/641443
; FILING DATE: 29-Apr-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/591925
; FILING DATE: 25-Jan-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Schwartz, Timothy R.
; REGISTRATION NUMBER: 32171
; REFERENCE/DOCKET NUMBER: P0989P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-7467
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 64 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: DNA (genomic)
; US-08-697-631A-12

Query Match          54.7%; Score 24.6; DB 3; Length 64;
Best Local Similarity 87.1%; Pred. No. 0.68;
Matches 27; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 9 tagtgatggtgatggtgatgaaagccaat 39
   | ||||| ||||| ||||| ||||| |||||
Db 17 TGTGTATGCTGATGCTGATGAATGCGATAT 47

RESULT 9
US-08-697-631A-13/C
; Sequence 13, Application US/08697631A
; Patent No. 5879673
; GENERAL INFORMATION:
; APPLICANT: Thomas, Griffith R
; TITLE OF INVENTION: No. 5879673el Administration of Thrombopoietin
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/697,631A
; FILING DATE: 28-Aug-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/641443
; FILING DATE: 29-Apr-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/591925
; FILING DATE: 25-Jan-1996
```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Schwartz, Timothy R.
; REGISTRATION NUMBER: 32171
; REFERENCE/DOCKET NUMBER: P0989P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-7467
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 65 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: DNA (genomic)
; US-08-697-631A-13

Query Match          54.7%; Score 24.6; DB 3; Length 65;
Best Local Similarity 87.1%; Pred. No. 0.68;
Matches 27; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 9 tagtgatggtgatggtgatgaaagccaat 39
   | ||||| ||||| ||||| ||||| |||||
Db 49 TGTGTATGCTGATGCTGATGAATGCGATAT 19

RESULT 10
US-08-697-631A-11/C
; Sequence 11, Application US/08697631A
; Patent No. 5879673
; GENERAL INFORMATION:
; APPLICANT: Thomas, Griffith R
; TITLE OF INVENTION: No. 5879673el Administration of Thrombopoietin
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/697,631A
; FILING DATE: 28-Aug-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/641443
; FILING DATE: 29-Apr-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/591925
; FILING DATE: 25-Jan-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Schwartz, Timothy R.
; REGISTRATION NUMBER: 32171
; REFERENCE/DOCKET NUMBER: P0989P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-7467
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 66 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: DNA (genomic)
; US-08-697-631A-11
```



;; TITLE OF INVENTION: Crosslinking  
;; NUMBER OF SEQUENCES: 35  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
;; STREET: Four Embarcadero Center, Suite 200  
;; CITY: San Francisco  
;; STATE: CA  
;; COUNTRY: US  
;; ZIP: 94111  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/642,246  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: ROWLAND, Bertram I  
;; REGISTRATION NUMBER: 20015  
;; REFERENCE/DOCKET NUMBER: A61127-1/BIR  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415-781-1989  
;; TELEFAX: 415-398-3249  
;; INFORMATION FOR SEQ ID NO: 21:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 32 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: other nucleic acid  
;; DESCRIPTION: /desc = "synthetic"  
US-08-642-246-21

Query Match 54.2%; Score 24.4; DB 5; Length 32;  
Best Local Similarity 96.2%; Pred. No. 0.71;  
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 gcttagtgatggtgatggtgaaga 31  
|||||  
DB 2 gcttagtgatggtgatggtgaaga 27

RESULT 14  
PCT-US96-06229-21  
;; Sequence 21, Application PC/TUS9606229  
;; GENERAL INFORMATION:  
;; APPLICANT: STEDRONSKY, Erwin R.  
;; APPLICANT: CAPELLLO, Joseph  
;; TITLE OF INVENTION: Tissue Adhesive Using Synthetic  
;; TITLE OF INVENTION: Crosslinking  
;; NUMBER OF SEQUENCES: 35  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
;; STREET: Four Embarcadero Center, Suite 200  
;; CITY: San Francisco  
;; STATE: CA  
;; COUNTRY: US  
;; ZIP: 94111  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US96/06229  
;; FILING DATE:  
;; CLASSIFICATION:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: ROWLAND, Bertram I  
;; REGISTRATION NUMBER: 20015

;; REFERENCE/DOCKET NUMBER: A61127-1/BIR  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415-781-1989  
;; TELEFAX: 415-398-3249  
;; INFORMATION FOR SEQ ID NO: 21:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 32 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: other nucleic acid  
;; DESCRIPTION: /desc = "synthetic"  
PCT-US96-06229-21

Query Match 54.2%; Score 24.4; DB 6; Length 32;  
Best Local Similarity 96.2%; Pred. No. 0.71;  
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 gcttagtgatggtgatggtgaaga 31  
|||||  
DB 2 gcttagtgatggtgatggtgaaga 27

RESULT 15  
US-08-707-237A-83  
;; Sequence 83, Application US/08707237A  
;; Patent No. 5830713  
;; GENERAL INFORMATION:  
;; APPLICANT: Ferrari, Franco A.  
;; APPLICANT: CAPELLLO, Joseph  
;; APPLICANT: Crissman, John W.  
;; APPLICANT: Dorman, Mary A.  
;; TITLE OF INVENTION: METHODS FOR PREPARING SYNTHETIC  
;; TITLE OF INVENTION: REPEITIVE DNA  
;; NUMBER OF SEQUENCES: 108  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
;; STREET: Four Embarcadero Center, Suite 3400  
;; CITY: San Francisco  
;; STATE: California  
;; COUNTRY: United States  
;; ZIP: 94111-4187  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/707,237A  
;; FILING DATE: 03-SEP-1996  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/175,155  
;; FILING DATE: 29-DEC-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/053,049  
;; FILING DATE: 22-APR-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/609,716  
;; FILING DATE: 06-NOV-1990  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/269,429  
;; FILING DATE: 09-NOV-1988  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/114,618  
;; FILING DATE: 29-OCT-1987  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 06/927,258  
;; FILING DATE: 04-NOV-1986  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Trecaulin, Richard F.  
;; REGISTRATION NUMBER: 31,801



REFERENCE/DOCKET NUMBER: A-55186-10/WH  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 781-1989  
 TELEFAX: (415) 398-3249  
 TELE: 910 277299  
 INFORMATION FOR SEQ ID NO: 83:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 31 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: unknown  
 TOPOLOGY: unknown  
 MOLECULE TYPE: DNA (genomic)  
 US-08-707-237A-83

Query Match 49.8%; Score 22.4; DB 3; Length 31;  
 Best Local Similarity 95.8%; Pred. No. 3.7;  
 Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 8 ttagtgatggtgatgaa 31  
 ||||||||||||||||  
 Db 3 ttagtgatggtgatgaa 26

Search completed: June 4, 2000, 16:09:46  
 Job time: 28065 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:05:41 ; Search time 1236.38 Seconds

(without alignments)  
-31.472 Million cell updates/sec

Title: US-09-164-714-17

Perfect score: 40  
Sequence: 1 gacggccggcgtgtatcaattgcatgacgaagcgtaagtt 40

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 356616

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Genembl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl1:\*  
10: gb\_pl2:\*  
11: gb\_pl3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vl:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_or:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vl:\*  
32: gb\_htg1:\*  
33: gb\_htg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pir4:\*  
41: gb\_htg3:\*  
42: gb\_htg4:\*  
43: gb\_htg5:\*  
44: gb\_htg6:\*

45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pl5:\*  
52: gb\_htg8:\*  
53: gb\_htg9:\*  
54: gb\_htg10:\*  
55: gb\_htg11:\*  
56: gb\_htg12:\*  
57: gb\_htg13:\*  
58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	16.8	42.0	81	13 HUMUT815A	L31216 Human STS U
2	16.6	41.5	86	9 D78279S03	D78281 Homo sapien
3	16.2	40.5	69	12 AF158139	AF158139 Mus muscu
4	16	40.0	20	5 A93559	A93559 Sequence 4
5	16	40.0	45	14 D64066	D64066 Synthetic c
6	16	40.0	62	10 S85457	S85457 Immunoglobu
7	15.6	39.0	60	10 S69473	S69473 Igh-Ig heav
8	15.4	38.5	60	10 S75386	S75386 Smu-S gamma
9	15.2	38.0	60	40 AF064896	AF064896 Homo sapl
10	15.2	38.0	91	12 RATTCBXL	M18850 Rat T-cell
11	15	37.5	50	12 MMU41929	U41929 Mus musculu
12	15	37.5	94	5 AR017623	AR017623 Sequence
13	14.8	37.0	62	5 A84780	A84780 Sequence 25
14	14.8	37.0	66	12 S73712	S73712 TCR V14J2.2
15	14.8	37.0	99	12 AF041885	AF041885 Mus muscu
16	14.6	36.5	35	5 A39021	A39021 Sequence 2
17	14.6	36.5	35	5 AR016008	AR016008 Sequence
18	14.6	36.5	39	24 E09619	E09619 PCR primer
19	14.6	36.5	60	12 M051GA4	M12445 Mouse Ig ge
20	14.6	36.5	74	9 HS33VB3	Z43302 H. sapiens T
21	14.6	36.5	88	34 TCUC39755	U39755 Trypanosoma
22	14.6	36.5	89	12 MMV5M204	Z12386 M. musculus
23	14.6	36.5	100	9 HS185B3F	Z60005 H. sapiens C
24	14.4	36.0	46	5 A93653	A93653 Sequence 5
25	14.4	36.0	57	13 HUMUT5347A	L30843 Human STS U
26	14.4	36.0	65	5 A87799	A87799 Sequence 18
27	14.4	36.0	65	5 A87800	A87800 Sequence 19
28	14.4	36.0	66	12 AF158167	AF158167 Mus muscu
29	14.4	36.0	78	10 HSU5079	U5079 Human 18cla
30	14.4	36.0	93	5 AR024220	AR024220 Sequence
31	14.4	36.0	94	13 HUMUT5143A	L31061 Human STS U
32	14.2	35.5	38	5 E05849	E05849 DNA sequenc
33	14.2	35.5	49	5 AR018160	AR018160 Sequence
34	14.2	35.5	45	5 I71071	I71071 Sequence 22
35	14.2	35.5	50	12 MMU41915	U41915 Mus musculu
36	14.2	35.5	60	5 I00798	I00798 Sequence 2
37	14.2	35.5	60	5 I00800	I00800 Sequence 4
38	14.2	35.5	73	5 AR012449	AR012449 Sequence
39	14.2	35.5	73	5 AR020277	AR020277 Sequence
40	14.2	35.5	73	5 I82623	I82623 Sequence 64
41	14.2	35.5	95	10 HSSRP2S05	U72200 Human pheno
42	14	35.0	22	5 AR049652	AR049652 Sequence
43	14	35.0	31	5 I28696	I28696 Sequence 18
44	14	35.0	37	5 A42247	A42247 Sequence 35
45	14	35.0	60	10 S69475	S69475 Igh-Ig heav

## ALIGNMENTS



Query Match 40.5%; Score 16.2; DB 12; Length 69;  
Best Local Similarity 64.9%; Pred. No. 1.1e+04;  
Matches 24; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 gagccggctgtatcataatgcatagcgagtaatt 40  
||||| | | | | | | | | | | | | | | | | |  
DB 23 GGCCCGGCGAGCTCTTGTGTGAAGGCTCAAGCT 59

RESULT 4  
A93559 20 bp DNA PAT 22-JAN-2000  
DEFINITION Sequence 4 from Patent WO9738105.  
ACCESSION A93559  
VERSION A93559.1 GI:6741764  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 20)  
AUTHORS Baier M. and Metzner K.  
TITLE GENOMIC NUCLEIC ACIDS, CDNA AND MRNA WHICH CODE FOR POLYPEPTIDES WITH IL-16 ACTIVITY, PROCESSES FOR THE PRODUCTION THEREOF AND THEIR USE  
JOURNAL Patent: WO 9738105-A 16-OCT-1997;  
FEATURES BUNDESRP DEUTSCHLAND (DE); BAIER MICHAEL (DE)  
source 1. .20  
location/Qualifiers  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 1 a 6 c 10 g 3 t  
LOCUS

ORIGIN

Query Match 40.0%; Score 16; DB 5; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.3e+04;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gagcgccggctgtgt 16  
||||| | | | | | | | | | | | | | | | | |  
DB 5 GACGGCCGGCGCTGT 20

RESULT 5  
D64066 45 bp DNA SYN 21-MAY-1999  
DEFINITION Synthetic construct DNA for TSS1-1 binding peptide, partial cds.  
ACCESSION D64066  
VERSION D64066.1 GI:1665732  
KEYWORDS TSS1-1 binding peptide.  
SOURCE synthetic construct DNA, clone:Phage 8.  
ORGANISM synthetic construct artificial sequence.  
REFERENCE 1 (sites)  
AUTHORS Sato,A., Ida,N., Fukuyama,M., Miwa,K., Kazami,J. and Nakamura,H.  
TITLE Identification from a phage display library of peptides that bind to toxic shock syndrome toxin-1 and that inhibit its binding to major histocompatibility complex (MHC) class II molecules  
JOURNAL Biochemistry 35 (32), 10441-10447 (1996)  
MEDLINE 9639315  
REFERENCE 2 (bases 1 to 45)  
AUTHORS Sato,A., Ida,N., Fukuyama,M., Miwa,K., Kazami,J. and Nakamura,H.  
TITLE Identification of Toxic Shock Syndrome Toxin-1 binding peptides that inhibit binding to MHC class II molecules from a phage display library  
JOURNAL Unpublished (1995)  
REFERENCE 3 (bases 1 to 45)  
AUTHORS Sato,A.  
TITLE Direct Submission  
JOURNAL Submitted (06-SEP-1995) to the DDBJ/EMBL/Genbank databases. Atsushi Sato, Biomolecular Engineering Research Institute, Department of Molecular Biology, Furuedai 6-2-3, Suita, Osaka 565-0874, Japan (Tel:06-872-8204, Fax:06-872-8219)

FEATURES Location/Qualifiers  
source 1. .45  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/clone="Phage 8"  
/note="synonym:artificial sequence"  
<1..>45  
/function="TSS1-1 binding"  
/codon\_start=1  
/transl\_table=11  
/evidence=experimental  
/product="TSS1-1 binding peptide"  
/protein\_id="BAI10941.1"  
/db\_xref="GI:1665733"  
/translation="WDAMTWNRSEFH"  
BASE COUNT 8 a 4 c 17 g 16 t  
ORIGIN

Query Match 40.0%; Score 16; DB 14; Length 45;  
Best Local Similarity 68.8%; Pred. No. 1.3e+04;  
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 9 gagctgtatcataatgcatagcgagtaatt 40  
||| | | | | | | | | | | | | | | | |  
DB 2 GGGATGCGATGTAFTGGATTTGGCGGAGTGTT 33

RESULT 6  
S85457/c 62 bp DNA PRI 07-MAY-1993  
LOCUS S85457/c  
DEFINITION immunoglobulin heavy chain complementarity determining region (CDR)-III {3' variable, diversity, and 5' joining regions} [human, B cell precursor acute lymphoblastic leukemia cells, Genomic, 62 nt].  
ACCESSION S85457  
VERSION S85457.1 GI:246644  
KEYWORDS  
SOURCE human B cell precursor acute lymphoblastic leukemia cells.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 62)  
AUTHORS Kiyoi,H., Naoe,T., Horibe,K. and Ohno,R.  
TITLE Characterization of the immunoglobulin heavy chain complementarity determining region (CDR)-III sequences from human B cell precursor acute lymphoblastic leukemia cells  
JOURNAL J. Clin. Invest. 89 (3), 739-746 (1992)  
MEDLINE 92176356  
REMARK Genbank staff at the National Library of Medicine created this entry [NCBI glibsq 85457] from the original journal article.  
This sequence comes from fig 1.  
location/Qualifiers  
1. .62  
source /organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1. .62  
gene /partial  
/gene="Immunoglobulin heavy chain complementarity determining region (CDR)-III"

BASE COUNT 10 a 17 c 23 g 12 t  
ORIGIN

Query Match 40.0%; Score 16; DB 10; Length 62;  
Best Local Similarity 68.8%; Pred. No. 1.3e+04;  
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 6 cccggctgtatcataatgcatagcgagtaaa 37  
||| | | | | | | | | | | | | | | | |  
DB 47 CCAGGGGTGCAACCACTTGTCTTAGAGGTTAA 16

RESULT 7  
LOCUS S69473 60 bp DNA PRI 23-SEP-1994  
DEFINITION IGH-Ig heavy chain [clone PSC1.1-11, mu to epsilon switch circle, recombination breakpoint] [human, B lymphocytes, Extrachromosomal, 60 nt].  
ACCESSION S69473 GI:546126  
VERSION S69473.1 GI:546126  
KEYWORDS  
SOURCE human B lymphocytes.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 60)  
AUTHORS Zhang, K., Mills, F.C. and Saxon, A.  
TITLE Switch circles from IL-4-directed epsilon class switching from human B lymphocytes. Evidence for direct, sequential, and multiple step sequential switch from mu to epsilon Ig heavy chain gene  
JOURNAL J. Immunol. 152 (7), 3427-3435 (1994)  
MEDLINE 94194102  
REMARK Genbank staff at the National Library of Medicine created this entry [NCBI g1bbsq 146011] from the original journal article. This sequence comes from Fig. 3A.  
FEATURES  
source 1..60  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..60  
/partial  
/gene="IGH"  
/note="Ig heavy chain"  
BASE COUNT 13 a 12 c 23 g 12 t  
ORIGIN  
Query Match 39.0%; Score 15.6; DB 10; Length 60;  
Best Local Similarity 81.8%; Pred. No. 2e+04;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 4 gggccgggtgtatcaattg 25  
||| ||||| ||| |||  
Db 33 GGCTGGCTGGAGTGAACCTGG 54  
RESULT 8  
LOCUS S75386 60 bp DNA PRI 27-JUL-1995  
DEFINITION Smu-S gamma 1 switch region: Ig VDJ [alternatively spliced, switch recombination sites] [human, isotype-switched B-lymphocytes, Genomic, 60 nt].  
ACCESSION S75386  
VERSION S75386.1 GI:914260  
KEYWORDS  
SOURCE human isotype-switched B-lymphocytes.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 60)  
AUTHORS Zhang, K., Chean, H.K. and Saxon, A.  
TITLE Secondary deletional recombination of rearranged switch region in Ig isotype-switched B cells. A mechanism for isotype stabilization  
JOURNAL J. Immunol. 154 (5), 2237-2247 (1995)  
MEDLINE 95173430  
REMARK Genbank staff at the National Library of Medicine created this entry [NCBI g1bbsq 160356] from the original journal article. This sequence comes from Fig. 4B.  
COMMENT Region: Smu-S gamma 1 switch region.  
FEATURES  
source 1..60  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..60  
/partial  
gene

BASE COUNT 10 a 13 c 24 g 13 t  
ORIGIN  
Query Match 38.5%; Score 15.4; DB 10; Length 60;  
Best Local Similarity 76.0%; Pred. No. 2.4e+04;  
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 4 gggccgggtgtatcaattgcat 28  
||| ||||| ||| |||  
Db 33 GGCTGGCTGGAGTGAACCTGGCTT 57  
RESULT 9  
LOCUS AF064896 60 bp mRNA PRI 03-MAR-1999  
DEFINITION Homo sapiens clone 11-4 immunoglobulin heavy chain alpha VDJ region mRNA, partial cds.  
ACCESSION AF064896  
VERSION AF064896.1 GI:4321882  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 60)  
AUTHORS Rowley, A.H., Eckerley, C.A., Shulman, S.T. and Baker, S.C.  
TITLE Evidence of an antigen-driven IgA immune response in the vascular wall in acute Kawasaki Syndrome  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 60)  
TITLE Rowley, A.H., Eckerley, C.A., Shulman, S.T. and Baker, S.C.  
JOURNAL Direct Submission  
TITLE Submitted (14-MAY-1998) Pediatrics, Loyola University, 2160 S First Avenue, Maywood, IL 60153, USA  
FEATURES  
source 1..60  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="11-4"  
/tissue-type="vascular"  
/note="from aorto-iliac junction of a 10-year-old child with fatal acute Kawasaki Syndrome"  
<1..>60  
/note="includes IGHV6 gene"  
/codon\_start=1  
/product="immunoglobulin heavy chain alpha VDJ region"  
/protein\_id="AAD15860.1"  
/db\_xref="GI:4321883"  
/translation="PDDTAVYYCARGVEGFDYWG"  
BASE COUNT 11 a 13 c 23 g 13 t  
ORIGIN  
Query Match 38.0%; Score 15.2; DB 40; Length 60;  
Best Local Similarity 63.9%; Pred. No. 2.9e+04;  
Matches 23; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 1 gacggccgggtgtatcaattgcatgagcggtga 36  
||| ||| ||| ||| ||| ||| |||  
Db 4 GACGACACGCGTGTGTACTGTGCACAGGGGGTA 39  
RESULT 10  
LOCUS RATTCBXL 91 bp mRNA ROD 27-APR-1993  
DEFINITION Rat T-cell receptor active beta-chain V-region (V-J-beta2.2) mRNA, partial cds, clone TRB89.  
ACCESSION M18850  
VERSION M18850.1 GI:207243  
KEYWORDS J-region; T-cell receptor; T-cell receptor beta-chain; V-region; processed gene.

SOURCE	Rat (6- to 8-week-old) thymocyte, cDNA to mRNA, clone TRB89.
ORGANISM	Rattus norvegicus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
AUTHORS	(bases 1 to 91)
TITLE	Morris, M., Barclay, A.N. and Williams, A.F.
JOURNAL	Analysis of T cell receptor beta chains in rat thymus, and rat
MEDLINE	Immunogenetics 27, 174-179 (1988)
FEATURES	88113841
SOURCE	1..91 Location/Qualifiers
BASE COUNT	19 a 19 c 33 g 20 t
ORIGIN	/organism="Rattus norvegicus" /db_xref="taxon:10116"
Query Match	38.0%; Score 15.2; DB 12; Length 91;
Best Local Similarity	63.9%; Pred. No. 2.9e+04;
Matches	23; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
Oy	5 gccccggctggtatcaattggcataggcggtgaagt 40
Db	45 GACCGGCGACTATCTTGTGTGAAGGCTCAAGCT 80
RESULT 11	
LOCUS	MMU41929 50 bp DNA ROD 05-JAN-1996
DEFINITION	Mus musculus recombination between immunoglobulin heavy chain and
ACCESSION	C-myc.
VERSION	U41929
KEYWORDS	U41929.1 GI:1147663
SOURCE	house mouse strain-BALB/cAn.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE	1 (bases 1 to 50)
JOURNAL	Muller, J.R.
FEATURES	Direct Submission
SOURCE	Submitted (05-DEC-1995) Jurgen R. Muller, Lab of Genetics, NIH/NCI,
BASE COUNT	Bldg. 37, Room 2B09, 37 Convent Dr., Bethesda, MD 20892-4255, USA
ORIGIN	Location/Qualifiers
1..50	/organism="Mus musculus"
/strain="BALB/cAn"	
/db_xref="taxon:10090"	
/chromosome="7(12;15)"	
/map="7(12f1;15d2)"	
/tissue_type="spleen 7 days post pristane"	
/dev_stage="7 days post pristane"	
10 a 21 c 13 g 6 t	
Query Match	37.5%; Score 15; DB 12; Length 50;
Best Local Similarity	67.7%; Pred. No. 3.5e+04;
Matches	21; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
Oy	4 gggccggcgtgatcaattggcataggcggtgaagt 34
Db	31 GTCCCGAGCTGGAATAGCTTGGGCTGGGCTG 1
RESULT 12	
LOCUS	AR017623 94 bp DNA PAT 05-DEC-1998
DEFINITION	Sequence 24 from patent US 5780228.
ACCESSION	AR017623
VERSION	AR017623.1 GI:3973226
KEYWORDS	

SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 94)
TITLE	Pama,D.H., Hicke,B., Bridonneau,P. and Gold,L.
JOURNAL	High affinity nucleic acid ligands to lectins
FEATURES	Patent: US 5780228-A 24-14-Jul-1998;
source	Location/Qualifiers 1..94
BASE COUNT	32 a 21 c 34 g 7 t
ORIGIN	/organism="unknown"
Query Match	37.5%; Score 15; DB 5; Length 94;
Best Local Similarity	61.5%; Pred.No. 3.5e+04;
Matches	24; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Oy	2 acggccgcggtcgtgatcaattgccatagcgcgtaagt 40   
Db	35 ATGCCCGGGCAGCAACTGCGTGCACAGCACAGCT 73 
RESULT	13
A84780	62 bp DNA PAT 21-JAN-2000
LOCUS	A84780
DEFINITION	Sequence 25 from Patent WO9844135.
ACCESSION	A84780
VERSION	A84780.1 GI:6733648
KEYWORDS	.
SOURCE	unidentified.
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 62)
AUTHORS	Hegemann,J. and Zimmermann,F.K.
TITLE	METHOD FOR SCREENING ANTIMYCOTICALLY ACTIVE SUBSTANCES
JOURNAL	Patent: WO 9844135-A 08-OCT-1998;
HEGEMANN JOHANNES (DE); ZIMMERMANN FRIEDRICH KARL (DE)	
FEATURES	Location/Qualifiers 1..62
source	/organism="unidentified" /db_xref="taxon:32644"
exon	1..62
BASE COUNT	17 a 15 c 12 g 18 t
ORIGIN	
Query Match	37.0%; Score 14.8; DB 5; Length 62;
Best Local Similarity	73.1%; Pred.No. 4.3e+04;
Matches	19; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Oy	14 ggtatcaattggcatagcgcgtaagt 39   
Db	30 GTTATCTATTAGCATYAGCCACTACT 55 
RESULT	14
S73712	66 bp mRNA ROD 02-MAR-1995
LOCUS	S73712
DEFINITION	TCR V4J2.2 beta "myelin basic protein-specific T-cell receptor
beta chain VDJ junction region [rats, Lewis, T cells, mRNA Partial,	
66 nt].	
ACCESSION	S73712
VERSION	S73712.1 GI:688024
KEYWORDS	.
SOURCE	Rattus sp. Lewis T cells.
ORGANISM	Rattus sp.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
TITLE	1 (bases 1 to 66) Sun,D., Shah,R. and Coleclough,C. Repertoire of rat MBP-reactive T cells: DNA sequencing analysis further demonstrates the clonal heterogeneity of rat T cells

reactive against encephalitogenic epitopes  
Cell. Immunol. 156 (2), 389-401 (1994)  
94298095  
JOURNAL MEDLINE  
Genbank staff at the National Library of Medicine created this  
entry [NCBI 91bbp 156543] from the original journal article.  
REMARK This sequence comes from Fig. 5.  
FEATURES  
Source  
1..66  
/organism="Rattus sp."  
/db\_xref="taxon:10118"  
1..66  
/partial  
/gene="TCR V14J2.2&br;"  
/note="myelin basic protein-specific T-cell receptor beta  
chain VDJ junction region"  
1..66  
/partial  
/gene="TCR V14J2.2&br;"  
/note="This sequence comes from Fig. 5; Protein sequence  
is in conflict with the conceptual translation;  
mismatches(15[E->Q],17[S->T],19[L->M])"  
/codon\_start=1  
/product="myelin basic protein-specific T-cell receptor  
beta chain VDJ junction region"  
/protein\_id="AAB31305.1"  
/db\_xref="GI:688025"  
/translation="CASHRRATGQLYFEGSGSKLTNVL"  
BASE COUNT 16 a 20 c 19 g 11 t  
ORIGIN  
Query Match 37.0%; Score 14.8; DB 12; Length 66;  
Best Local Similarity 64.7%; Pred. No. 4.3e+04;  
Matches 22; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
QY 7 ccggcgtgatacattgacatgacggtgaactt 40  
||||| | | | | | | | | | | | | | | | | |  
DB 23 CCGGGCAGCTTACTTGGTGAAAGCCTTAAGAT 56  
RESULT 15  
AF041885 99 bp mRNA ROD 20-JAN-1998  
LOCUS Mus musculus uninfected mouse C seq 1, T cell receptor beta chain  
DEFINITION mRNA, partial cds.  
ACCESSION AF041885  
VERSION AF041885  
KEYWORDS AF041885.1 GI:2795992  
SOURCE  
house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 99)  
AUTHORS Gomez,F.J., Cain,J.A., Gibbons,R., Allendeofer,R. and Deepe,G.S.  
TITLE V beta 4 T cells exert protection in pulmonary experimental  
histoplasmosis  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 99)  
AUTHORS Gomez,F.J., Cain,J.A., Gibbons,R., Allendeofer,R. and Deepe,G.S.  
TITLE Direct Submission  
JOURNAL Submitted (09-JAN-1998) Internal Medicine, Division of Infectious  
Diseases, University of Cincinnati, 231 Bethesda Ave. ML 560,  
Cincinnati, OH 45267-0560, USA  
FEATURES  
Source  
1..99  
/organism="Mus musculus"  
/strain="C57BL/6"  
/isolate="uninfected control, mouse C seq 1"  
/db\_xref="taxon:10090"  
/tissue\_type="lung"  
/note="type: jbeta2.2"  
<1..15  
/standard\_name="V beta 4"

CDS  
<1..>99  
/note="CDR3 region"  
/codon\_start=1  
/product="T cell receptor beta chain"  
/protein\_id="AAB97192.1"  
/db\_xref="GI:2795993"  
/translation="CASSQGLGNTGQLYFEGSGSKLTVLEDLRNVT"  
16..27  
/standard\_name="N sequences and D beta 2"  
28..76  
/standard\_name="J beta 2.2"  
77..>99  
/standard\_name="C beta"  
BASE COUNT 24 a 23 c 33 g 19 t  
ORIGIN  
Query Match 37.0%; Score 14.8; DB 12; Length 99;  
Best Local Similarity 64.7%; Pred. No. 4.3e+04;  
Matches 22; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
QY 7 ccggcgtgatacattgacatgacggtgaactt 40  
||||| | | | | | | | | | | | | | | | | |  
DB 32 CCGGGCAGCTTACTTGGTGAAAGCCTTAAGCT 65  
Search completed: June 4, 2000, 16:05:43  
Job time: 27895 sec





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:24:18 ; Search time 322.35 Seconds

(without alignments)  
31.046 Million cell updates/sec

Title: US-09-164-714-17

Perfect score: 40  
Sequence: 1 gagcgccggcgtgtatcaattgcatagcgagtaagt 40

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 431286

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.8	42.0	99	1	V39623
2	16	40.0	28	1	Q93906
3	16	40.0	48	1	V10672
4	16	40.0	60	1	V40188
5	16	40.0	83	1	V99110
6	15.6	39.0	51	1	X15369
7	15.4	38.5	30	1	T59426
8	15.4	38.5	54	1	Q3546
9	15	37.5	58	1	V28977
10	15	37.5	94	1	T57700
11	14.8	37.0	62	1	V64262
12	14.6	36.5	39	1	T02866
13	14.6	36.5	87	1	T36186
14	14.6	36.5	87	1	T37293
15	14.6	36.5	87	1	T51211
16	14.6	36.5	87	1	T68704
17	14.4	36.0	61	1	T23877
18	14.4	36.0	63	1	V76296
19	14.4	36.0	97	1	T58032
20	14.2	35.5	38	1	O64620
21	14.2	35.5	47	1	V07587
22	14.2	35.5	48	1	Q34008
23	14.2	35.5	60	1	T18142
24	14.2	35.5	73	1	T71390
25	14	35.0	22	1	T01364
26	14	35.0	22	1	V12409
27	14	35.0	22	1	V33803
28	14	35.0	31	1	T46172
29	14	35.0	37	1	T34354
30	14	35.0	55	1	Q33999
31	14	35.0	61	1	X33461
32	14	35.0	63	1	T04873
33	14	35.0	96	1	T58022
34	13.8	34.5	18	1	Q12438

35	13.8	34.5	25	1	Q12443	Primer HPUS-1 for
36	13.8	34.5	36	1	X04271	ITV linker regions
37	13.8	34.5	55	1	T59173	DNA polymerase IIg
38	13.8	34.5	60	1	V40207	Escherichia coli B
39	13.8	34.5	77	1	T88485	Secretory phosphol
40	13.8	34.5	94	1	Q89695	Vector back-end ol
41	13.8	34.5	94	1	O89696	Dengue virus type
42	13.6	34.0	36	1	V29556	Dengue virus type
43	13.6	34.0	55	1	T34005	Primer for amplifi
44	13.6	34.0	66	1	Q36624	Fragment of homolo
45	13.6	34.0	66	1	T00547	Equine herpesvirus

## ALIGNMENTS

RESULT 1  
V39623 V39623 standard: cDNA; 99 BP.  
AC V39623;  
DE 28-SEP-1998 (first entry)  
KW pfrc1 DNA oligonucleotide seq ID NO:141.  
KW Mass spectrometry; diagnosis; detection; biological sample; infection;  
KW genetic disease; chromosomal abnormality; identification; heredity;  
KW pathogenic organism; telomerase activity; oncogene mutation;  
KW cancer-specific sequence; primer; ss.  
OS Synthetic.  
PN WO9820166-A2.  
PD 14-MAY-1998.  
PF 06-NOV-1997; U20444.  
PR 08-OCT-1997; US-947801.  
PR 06-NOV-1996; US-744481.  
PR 06-NOV-1996; US-744590.  
PR 06-NOV-1996; US-746036.  
PR 06-NOV-1996; US-746055.  
PR 23-JAN-1997; US-786988.  
PR 23-JAN-1997; US-787639.  
PR 19-SEP-1997; US-933792.  
PA (SEQU-) SEQUENOM INC.  
PI Braum A, Damhofer-Demar B, Fu D, Higgins GS, Jurinke C,  
PI Koster H, Little DP, Lough DM, Slegert CW, Tang K,  
PI Van Den Boom D, Xiang G;  
DR WPI, 98-286975/25.  
PT Sequencing nucleic acid by mass spectrometric analysis - for  
PT detecting nucleic acids, telomerase activity, oncogene mutations, or  
PT cancer-specific sequences, for diagnosis of disease  
PS Example 8; Page 283; 478pp; English.  
CC A process has been developed for determining the sequence of a target  
CC nucleic acid. The process comprises: (i) generating at least two  
CC fragments (F) from the target nucleic acid; and (ii) analysing F by  
CC mass spectrometry (MS). The sequences in V39483 to V39592 are  
CC specifically claimed primers for use in the mass spectrometric analysis  
CC of the above process. The process is used to detect genetic diseases  
CC (e.g. haemophilia, thalassemia, Duchenne muscular dystrophy, Alzheimer's  
CC disease, cystic fibrosis and many others) or chromosomal abnormalities  
CC (or predisposition); infections and cancers; also for establishing  
CC identity and heredity. Particular applications are diagnosis of  
CC neuroblastoma, detecting telomerase, determining family relationships  
CC and HLA compatibility, and in genetic fingerprinting. Compared with  
CC known methods using MS, this process requires fewer specific reagents  
CC and is better suited to automation. Extended primers are shorter;  
CC primer annealing is more efficient and the process allows detection of  
CC many sequences simultaneously. The present sequence represents an  
CC oligonucleotide used in an example from the present invention.  
SO Sequence 99 BP; 24 A; 34 C; 20 G; 21 T;

Query Match 42.0%; Score 16.8; DB 1; Length 99;  
Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
Matches 24; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

5 gcccgccgtgtatcaattgcatagcgagtaagt 40  
||| ||||| ||||| ||| ||| |||

Db 59 GCATGGCTGCAGCCCAATTGCATGGCCGTCGTTT 94

RESULT 2

ID 093906/c

AC 093906 standard; DNA: 28 BP.

DT 15-JAN-1996 (first entry)

DE Human mismatch repair pathway gene MSH2 exon primer, 16066.

KW Mismatch repair; MSH2; primer; identification; defect; alteration;

KW cancer; tumour; vaccine; ss.

OS Homo sapiens.

PN W09514085-A2.

PD 26-MAY-1995.

PF 17-NOV-1994; U13385.

PR 17-NOV-1993; US-154792.

PR 07-DEC-1993; US-163449.

PR 13-JUN-1994; US-259310.

PA (UYVE-) UNIV VERMONT & STATE AGRIC COLLEGE.

PA (DAND ) DANA FARMER CANCER INST.

PI Fishel R, Kolodner RD, Reenan RAG;

DR WPI: 95-200377/26.

PT Determining alteration in human mismatch repair pathways - used in

PT the diagnosis, prognosis and therapy of cancers and in screening

PT assays

PS Claim 30: Page 176; 256pp; English.

CC 093906 is an oligonucleotide primer used to amplify an exon

CC of the human mismatch repair pathway gene MSH2. Defects or

CC alterations in such a gene result in the accumulation of unstable

CC repeated DNA sequences, a feature of a number of different cancers.

CC The identification of a defect in the mismatch repair pathway can

CC be diagnostic of a predisposition to cancer and prognostic for a

CC particular mammalian cancer e.g colorectal, ovarian, endometrial

CC (uterine), renal, bladder, skin, rectal and bowel. The nucleotide

CC sequences and polypeptides of the hMSH2 gene may also be used for

CC therapy and in vaccines.

SQ Sequence 28 BP: 8 A; 8 C; 5 G; 7 T;

Query Match 40.0%; Score 16; DB 1; Length 28;

Best Local Similarity 79.2%; Pred. No. 2.1e+02;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 9 ggcctgcatcattgcatgacgc 32

Db 24 GGACGTTCATGCATGACATGACGC 1

RESULT 3

ID V10672 standard; cDNA: 48 BP.

AC V10672;

DT 02-JUL-1998 (first entry)

DE Mouse Fas ligand-like protein oligonucleotide #3.

KW Fas antigen; ligand-like; apoptosis-inducing activity; therapeutic;

KW prophylactic; cancer; viral infection; hepatitis; nephritis; influenza;

KW bone disease; atherosclerosis; inflammation; neurodegenerative disease;

KW ischaemic disease; autoimmune disease; diabetes; cardiomyopathy;

KW ulcerative colitis; primer; ss.

OS Synthetic.

OS Mus sp.

PN W09803648-A1.

PD 29-JAN-1998.

PF 17-JUL-1997; J02480.

PR 31-JAN-1997; JP-019330.

PR 19-JUL-1996; JP-191204.

PR 09-AUG-1996; JP-211695.

PA (TAKE ) TAKEDA CHEM IND LTD.

PI Hikiuchi Y, Nishi K, Shintani Y;

DR WPI: 98-120773/11.

PT Fas ligand-like protein with apoptosis inducing activity - useful as

PT therapeutic or prophylactic agent for cancer, viral infection,

PT Helicobacter pylori infection, invasive staphylococci, hepatitis,

PT nephritis, bone disease, atherosclerosis or pain

PS Example 2; Page 152; 120pp; English.

CC Oligonucleotides V10670-V10674 are used in a method to clone a novel

CC mouse Fas ligand-like protein with apoptosis inducing activity. This

CC protein can be used as a therapeutic and/or prophylactic agent for

CC cancer, viral infection, Helicobacter pylori infection, invasive

CC staphylococci, hepatitis, nephritis, bone disease, atherosclerosis or

CC pain. The protein can also be used in a method to treat acquired immune

CC deficiency syndrome (AIDS), joint tissues in rheumatism, inflammation,

CC hepatitis, neurodegenerative disease due to osteomyelodysplasia,

CC ischaemic disease, autoimmune disease, diabetes, diabetic complications,

CC cardiomyopathy, influenza, glomerulonephritis or ulcerative colitis.

SQ Sequence 48 BP: 9 A; 13 C; 17 G; 9 T;

Query Match 40.0%; Score 16; DB 1; Length 48;

Best Local Similarity 100.0%; Pred. No. 2.3e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gacggccgcggcgtgt 16

Db 33 GACGCCGCCGGCTGCT 48

RESULT 4

ID V40188 standard; DNA: 60 BP.

AC V40188;

DT 08-OCT-1998 (first entry)

DE Escherichia coli B1AP gene codon usage library construction oligo 9.

KW Escherichia coli; B1AP; codon usage library; construction; evolution;

KW protein engineering; industrial; pharmaceutical; recombination;

KW selection; promoter; ss.

OS Synthetic.

OS Escherichia coli.

PN W09827230-A1.

PD 25-JUN-1998.

PF 17-DEC-1997; U24239.

PR 18-DEC-1996; US-769062.

PA (MAXY-) MAXYGEN INC.

PI Patten PA, Stemmer WPC;

DR WPI: 98-362797/31.

PT Evolving proteins encoded by DNA substrates - by generating a

PT library of recombinant DNA molecules and screening for improved

PT properties such as activity, expression, solubility or specificity

PS Example; Page 61; 122pp; English.

CC A method has been developed for evolving a protein encoded by a DNA

CC substrate molecule. The method comprises: (a) digesting at least a first

CC and second DNA substrate molecule with restriction endonuclease, where

CC the first and second substrate molecules differ from each other in at

CC least one nucleotide; (b) ligating the mixture to generate a library of

CC recombinant DNA molecules; (c) screening or selecting the products of

CC (b) for a desired property; and (d) recovering a recombinant DNA

CC substrate molecule encoding an evolved protein. The method can be used

CC for evolving proteins to provide improved properties such as expression,

CC folding, solubility, activity, substrate specificity, shelf-life and

CC serum stability or circulation half-life. V40180 to V40235 represent

CC oligonucleotides used in the construction of a codon usage library in an

CC example from the present invention, describing a strategy to evolve B1AP

CC from Escherichia coli.

SQ Sequence 60 BP: 9 A; 12 C; 22 G; 13 T;

Query Match 40.0%; Score 16; DB 1; Length 60;

Best Local Similarity 64.3%; Pred. No. 2.4e+02;

Matches 18; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 7 ccggcgtgttcattgcatgacgcg 34

Db 44 CCAGRCRATATCCATGCAATGCCGCG 17

RESULT 5

V99110  
ID V99110 standard; DNA; 83 BP.  
AC V99110:  
DE 09-MAR-1999 (first entry)  
KW DNA methyltransferase genomic fragment.  
KW DNA methyltransferase; DNA Methylase; antisense oligonucleotide; human;  
KW cellular growth; tumour growth inhibition; silenced gene activation;  
KW beta thalassemia; sickle cell anemia; ss.  
OS Homo sapiens.  
PD WO9854313-A2.  
PN 03-DEC-1998.  
PF 29-MAY-1998; IB1107.  
PR 17-DEC-1997; US-069865.  
PR 30-MAY-1997; US-866340.  
PA (UYMC-) UNIV MCGILL.  
PI Bigey P, Ramchandani S, Szyf M;  
DR WPI: 99-059833/05.  
PT New DNA methyltransferase nucleotide sequences - used particularly  
PT to develop antisense oligonucleotides for diagnostic and therapeutic  
PT purposes, particularly for inhibiting tumour growth  
PS Claim 1, Page 52; 108pp; English.  
CC V99091-128 represent DNA methyltransferase (DNA Methylase) genomic  
CC fragments. The sequences together sequentially produce the full  
CC length sequence V99128. The sequences can be used to derive antisense  
CC oligonucleotides which inhibit DNA Methylase expression. The antisense  
CC oligonucleotides can be used in investigating the role of DNA Methylase  
CC in cellular growth. They can be administered at different points in  
CC the cell cycle, or in conjunction with promoters or inhibitors of cell  
CC growth to determine the role of DNA Methylase in the growth of the cell  
CC type of interest. The antisense oligonucleotides can also be used for  
CC inhibiting tumour growth in a mammal, or to activate silenced genes to  
CC provide a missing gene function. This ameliorates disease symptoms,  
CC e.g. in beta thalassemia and sickle cell anemia. The antisense  
CC oligonucleotides can also be used as an analytical and diagnostic tools  
CC and a potentiators of transgenic plant and animal studies.  
SQ Sequence 83 BP; 24 A; 18 C; 21 G; 20 T;

Query Match 40.0%; Score 16; DB 1; Length 83;  
Best Local Similarity 68.8%; Pred. No. 2.5e+02;  
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 2 acggccggcgtgataatgcatagcg 33  
DB 25 ACAGCCCCGTTGCTCTACTTAGAAAGGGG 56

RESULT 6  
X15369  
ID X15369 standard; DNA; 51 BP.  
AC X15369;  
DT 22-JUN-1999 (first entry)  
DE HSV-1; thymidine kinase gene primer MB127.  
KW HSV-1; thymidine kinase; mutation; DRH nucleoside binding site; enzyme;  
KW pathogen; tumour; hyperkeratosis; psoriasis; prostate hypertrophy;  
KW hyperthyroidism; endocrinopathy; autoimmune disease; allergy; restenosis;  
KW viral disease; AIDS; hepatitis; parasite; bacterial infection; ss.  
OS Synthetic.  
OS Herpes simplex virus.  
PN US5877010-A.  
PD 02-MAR-1999.  
PF 02-MAY-1995; 432871.  
PR 02-MAY-1995; US-432871.  
PR 02-MAY-1994; US-237592.  
PA (UNIV ) UNIV WASHINGTON.  
PI Black ME, Loeb LA;  
DR WPI: 99-189650/16.  
PT New Herpesviridae thymidine kinase mutant nucleic acids - used to  
PT develop products for treating e.g. tumours, autoimmune diseases,  
PT allergies, restenosis or viral, bacterial or parasitic diseases  
PS Example 8; Column 40; 72pp; English.  
CC This sequence represents a primer used to construct a mutated herpes  
CC simplex virus type 1 (HSV-1) thymidine kinase (TK) gene with mutations

CC In codons 159-161 and 168-170. The invention relates to the generation  
CC of novel HSV-1 TK gene with a mutation upstream, within or downstream  
CC from a DRH nucleoside binding site. The TK enzymes can be used for  
CC inhibiting pathogenic agents, e.g. tumours, hyperkeratosis, psoriasis,  
CC prostate hypertrophy, hyperthyroidism, endocrinopathies, autoimmune  
CC diseases, allergies, restenosis, viral diseases such as AIDS, hepatitis,  
CC intracellular parasitic diseases or bacterial infection.  
SQ Sequence 51 BP; 8 A; 9 C; 20 G; 5 T;

Query Match 39.0%; Score 15.6; DB 1; Length 51;  
Best Local Similarity 53.8%; Pred. No. 3.5e+02;  
Matches 21; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

OY 2 acggccggcgtgataatgcatagcggaagtc 40  
DB 8 ACCGCGCGCGCGGTACGANNNNNNNGCGGATGAT 46

RESULT 7  
T59426  
ID T59426 standard; DNA; 30 BP.  
AC T59426;  
DT 26-MAY-1997 (first entry)  
DE Human papillomavirus 6b early ORF E2 gene forward primer.  
KW Human papillomavirus; HPV; polyprotein; E2 gene;  
KW early open reading frame; vaccine; wart; condylomata acuminata;  
KW cervical cancer; polymerase chain reaction; PCR; primer; ss.  
OS Synthetic.  
PN WO9705164-A1.  
PD 13-FEB-1997.  
PF 26-JUL-1996; A00473.  
PR 27-JUL-1995; A0-004439.  
PA (CSLC-) CSL LTD.  
PA (UYOU ) UNIV QUEENSLAND.  
PI Cox JC, Edwards SJ, Frazer I, Margetis MB, McMillan NAJ;  
PI Moloney MBH, Webb EA, Williams MP;  
DR WPI: 97-145619/13.  
PT Papilloma virus early open reading frame polyprotein constructs -  
PT useful for treating, e.g. genital warts and cervical cancer  
PS Example 1; Page 18; 62pp; English.  
CC PCR primers (T59426-27) were designed to amplify the E2 gene  
CC of human papillomavirus 6b (HPV6b), incorporating a 5' SmaI site  
CC and 3' XbaI, NcoI, KpnI and SacI sites, and using the entire genome  
CC of HPV6b in pBR322 as template. The HPV6b E2b gene was similarly  
CC amplified (see also T59428-29). The amplified genes were similarly  
CC into an HPV6b E6/E4 cassette (see also T59412-15) in pSP70. The  
CC resulting cassettes (see also T59430-35) can be cloned into  
CC pCEX-STOP to allow prodn. of HPV6b polyproteins (W01808-13) in E.  
CC coli for use as vaccines.  
SQ Sequence 30 BP; 8 A; 9 C; 11 G; 2 T;

Query Match 38.5%; Score 15.4; DB 1; Length 30;  
Best Local Similarity 76.0%; Pred. No. 3.8e+02;  
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 3 cggccggcgtgataatgca 27  
DB 2 CGCCCGCGGATGAGCAATAGCCA 26

RESULT 8  
Q33546  
ID Q33546 standard; DNA; 54 BP.  
AC Q33546;  
DT 02-FEB-1993 (first entry)  
DE Downstream sequence of microsatellite from clone AG1A233.  
KW PCR; selection; primers; OPTIPRIM; breeding; cattle; parentage;  
KW genetic mapping; traits; amplification; ss.  
OS Bos taurus.  
PN WO9213102-A.  
PD 06-AUG-1992.



```
RESULT 11
V64262
ID V64262 standard; DNA; 62 BP.
AC V64262;
DE 25-JAN-1999 (first entry)
DE PCR primer YPL242c-S2.
KM Antimycotic agent; target; medicine; infection; veterinary; fungicide;
KW immunodepression; preservative; food industry; fungi; PCR primer; ss.
OS Synthetic.
PN WO9844135-A2.
PD 08-OCT-1998.
PE 02-APR-1998: E01904.
PR 02-APR-1997: DE-013572.
PA (FARN ) HOECHST AG.
PI Erlan K, Feldmann H, Hegemann J, Hinnen A, Koetter P,
PI Kramer W, Munder T, Rose M, Schuster T, Zimmermann FK;
DR WPI: 98-557125/47.
PT Identification of antimycotic agents using essential fungal proteins
PT or genes as targets - useful, e.g. for potential clinical, human or
PT veterinary medicine, for treatment of existing infections and for
PT prevention of these in immune depressed subjects
PS Disclosure; Page 38; 76pp; German.
CC V64259-V64274 are primers used in a method for the identification of
CC antimycotic agents using as a target a nucleic acid which controls an
CC essential protein of Saccharomyces cerevisiae or from other species of
CC Mycetes. Such agents are potentially useful clinically, in human or
CC veterinary medicine, for treating existing infections and for preventing
CC them in immune-depressed subjects (those with human immune deficiency
CC virus infection or diabetes), also as fungicides and preservatives for
CC foods and body care products. The agents are used to identify equivalent
CC genes in other fungi, specifically Candida albicans or Aspergillus
CC fumigatus, and equivalent human, animal and plant genes, and also for
CC identification of antimycotic agents.
SQ Sequence 62 BP; 17 A; 15 C; 12 G; 18 T;

Query Match 37.0%; Score 14.8; DB 1; Length 62;
Best Local Similarity 73.1%; Pred. No. 8e+02;
Matches 19; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 14 ggtacattgcatagcggttaagt 39
DB 30 gttatctattagcattagccactagtt 55

RESULT 12
T02866
ID T02866 standard; DNA; 39 BP.
AC T02866;
DE 14-MAR-1996 (first entry)
DE Fungal 18S rRNA encoding DNA amplification/sequencing primer, UFI.
KM Polymerase chain reaction; primer: ribosomal RNA; amplification;
KW sequencing; Matsutake mushroom; ss.
OS Agaricus bisporus.
PN J07177889-A.
PD 18-JUL-1995.
PE 22-DEC-1993: 346106.
PR 22-DEC-1993: JP-346106.
PA (RIKA ) RIKAGAKU KENKYUSHO.
DR WPI: 95-279918/37.
PT Oligonucleotide primer comprising amplification and sequencing
PT portions - useful for determination of fungal DNA sequences by PCR
PT amplification
PS Example 2: Page 6-7; 8pp; Japanese.
CC T02855-T02860 are amplification primers for DNA coding for
CC fungus-derived 18S rRNA. These primers may be bound at the 5' end
CC to the 3' end of a sequencing primer (T02861-T02863). The
CC resulting oligonucleotide primers comprising amplification and
CC sequencing portions (T02864-T02869). These primers are useful for
CC the determination of the base sequences of fungi.
SQ Sequence 39 BP; 12 A; 9 C; 9 G; 9 T;
```

```
Query Match 36.5%; Score 14.6; DB 1; Length 39;
Best Local Similarity 81.0%; Pred. No. 9e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gacggccggcgctgatatcgata 21
DB 9 GACGGCCAGTCTGTTGCTCA 29

RESULT 13
T36186/c
ID T36186 standard; DNA; 87 BP.
AC T36186;
DE 28-OCT-1996 (first entry)
DE Tagged protein DET1-DET2-spacer-SH2 DNA 5' primer.
KW Bone resorption disease; osteoporosis; src SH2 domain antagonist;
KW src homology 2 domain; polymerase chain reaction; PCR; primer; ss.
OS Synthetic.
PN EP-727211-A1.
PD 21-AUG-1996.
PE 07-FEB-1996; 200270.
PR 10-FEB-1995; US-386381.
PR 07-MAR-1995; US-400220.
PR 30-JUN-1995; US-497357.
PR 11-OCT-1995; US-541080.
PR 29-DEC-1995; US-580868.
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PI Dunnington DJ;
DR WPI: 96-372674/38.
PT Use of selective src SH2 domain ligand - to prepare medicament for
PT treating bone resorption disease
PS Example 11; Page 17; 47pp; English.
CC A PCR primer pair (T36186 and T36187) was used to amplify a DNA
CC sequence encoding the tagged protein DET1-DET2-spacer-SH2, with
CC insertion of 5' NdeI and BamHI sites and 3' XbaI site, using a cDNA
CC clone contg. the chicken src gene as template. In the tagged
CC protein, DET1 is an epitope (W02128) from HIV-1 gp120 or gp160.
CC DET2 is a hexahistidine tag and SH2 is the chicken src SH2 domain.
CC DNA encoding DET1-DET2-spacer-SH2 was used to produce vectors
CC encoding fusion proteins (see also W02119-21) useful in binding
CC assays to determine the specificity of cpts. to inhibit SH2 domains;
CC cpts. that selectively inhibit the human src SH2 domain can be used
CC to treat bone resorption diseases such as osteoporosis.
SQ Sequence 87 BP; 20 A; 28 C; 20 G; 19 T;

Query Match 36.5%; Score 14.6; DB 1; Length 87;
Best Local Similarity 69.0%; Pred. No. 1e+03;
Matches 20; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 gacggccggcgctgatatcgata 29
DB 43 GACGGCCCGGGCCACGCTGAATACGAATA 15

RESULT 14
T37293/c
ID T37293 standard; cDNA; 87 BP.
AC T37293;
DE 30-APR-1997 (first entry)
DE Chicken Src primer #1.
KM Polymerase chain reaction; PCR; amplify; primer; chicken; src;
KW SH2 domain; DET1; DET2; erythropoiesis; anaemia; haematopoiesis;
KW antagonist; ss.
OS Synthetic.
PN EP-728482-A2.
PD 28-AUG-1996.
PE 07-FEB-1996; 200269.
PR 10-FEB-1995; US-386381.
PR 07-MAR-1995; US-400220.
PR 30-JUN-1995; US-497357.
PR 11-OCT-1995; US-540680.
```

PR 29-DEC-1995: US-581089.  
PA (SMK ) SMITHKLINE BEECHAM CORP.  
PI Dunnington DJ;  
DR WPI: 96-386024/39.  
PT Use of selective antagonist of haematopoietic acid phosphatase SH2  
PT domain - with no significant affinity for other SH2 domains, to  
PT increase erythropoiesis and haematopoiesis, esp. for treatment of  
PT anaemia  
PS Example 1: Page 17; 46pp; English.  
CC The sequences given in T37293-94 are primers which were used in the  
CC cloning of the chicken src SH2 domain containing tags DRT1 and DRT2.  
CC The amplified sequence encodes a product which was used in the  
CC isolation of a compound for improving erythropoiesis. The compound  
CC may be used for the treatment of anaemia or to enhance haematopoiesis.  
CC The isolated compound antagonises the hcp SH2 domain without side  
CC effects caused by non-specific inhibition of other SH2 domains.  
SQ Sequence 87 BP; 20 A; 28 C; 20 G; 19 T;

Query Match 36.5%; Score 14.6; DB 1; Length 87;  
Best Local Similarity 69.0%; Pred. No. 1e+03;  
Matches 20; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 gacggccggcgctgataatggcata 29  
|||||  
DB 43 GACGGCCGGCGCCAGCTGAATACGATA 15

RESULT 15  
T51211/c  
ID T51211 standard; DNA; 87 BP.  
AC T51211;  
DE 20-OCT-1997 (first entry)  
DE Primer #1 for chicken src SH2 domain.  
KW HIV-1; gp120; defined epitope tag; DRT1; envelope protein; human; Stat 6;  
KW signal transduction and activation of transcription; Src homology 2; PCR;  
KW signaling molecule; protein tyrosine kinase; oncogenic protein; hcp SH2;  
KW Grb2 SH2; allergic reaction; erythrocyte production; inhibitor; p85 SH2;  
KW asthma; allergic rhinitis; atopic dermatitis; IGE receptor; SH-PTP2 SH2;  
KW Interleukin-4; IL-4; SH2 domain; Stat5 SH2; src SH2; lck SH2; fyn SH2;  
KW IL-13; therapy; primer; amplify; polymerase chain reaction; ss.  
OS Synthetic.  
PN WO9702023-A1.  
PD 23-JAN-1997.  
PF 28-JUN-1996; U11074.  
PR 08-FEB-1996; US-598716.  
PR 30-JUN-1995; US-497357.  
PA (SMK ) SMITHKLINE BEECHAM CORP.  
PI Dunnington DJ;  
PI WPI: 97-108735/10.  
PT Treating allergies with specific inhibitor of human Stat 6 SH2  
PT domain - having very low binding affinity to panel of other SH2  
PT domains so free of side effects, specifically for asthma and  
PT allergic rhinitis  
PS Example 11: Page 30; 88pp; English.  
CC T51211-T51220 represent amplification primers used in the construction of  
CC fusion proteins containing the src homology 2 (SH2) domain. The fusion  
CC proteins are used to identify a compound that targets the human Stat  
CC (signal transduction and activation of transcription) 6 SH2 domain. The  
CC identified compounds have a binding affinity for Stat 6 over 50  
CC (preferably 100) times higher than its affinity for the human Stat5 SH2  
CC domain. The compound has an affinity for hcp SH2, SH-PTP2 SH2, p85 SH2,  
CC Grb2 SH2, src SH2, lck SH2 or fyn SH2 of more than 50 (preferably 100)  
CC times lower than its affinity for Stat 6 SH2. SH2 domains are conserved  
CC non-catalytic sequences found in a variety of signalling molecules, such  
CC as non-receptor protein tyrosine kinases, and in oncogenic proteins. The  
CC compounds identified using the fusion proteins are used as the  
CC administered compound in the method of the invention for treating  
CC allergic reactions. Administration of the compound avoids the side  
CC effects (e.g. reduced erythrocyte production) associated with  
CC non-selective inhibition of SH2 domains. Selective compounds can be  
CC identified in competitive binding assays using only a small subset (the  
CC domains specified above) of SH2 domains rather than all 60 known domains.

CC The method can be used for the treatment of asthma and allergic rhinitis,  
CC but can also be used to treat atopic dermatitis. Inhibition of the human  
CC Stat 6 SH2 domain blocks up-regulation of the IGE receptor mediated by  
CC Interleukin-4 (IL-4) or IL-13.  
SQ Sequence 87 BP; 20 A; 28 C; 20 G; 19 T;

Query Match 36.5%; Score 14.6; DB 1; Length 87;  
Best Local Similarity 69.0%; Pred. No. 1e+03;  
Matches 20; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 gacggccggcgctgataatggcata 29  
|||||  
DB 43 GACGGCCGGCGCCAGCTGAATACGATA 15

Search completed: June 4, 2000, 16:24:20  
Job time: 28916 sec



Tue Jun 6 08:33:43 2000

us-09-164-714-17.rng

Page 7

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 13:53:29 ; Search time 4521.53 Seconds  
(without alignments)  
35.857 Million cell updates/sec

Title: US-09-164-714-17  
Perfect score: 40  
Sequence: 1 gacggccggcgctgtatcaattgcatagcggttaagt 40

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 119968

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: gb\_est1: \*  
21: gb\_est2: \*  
22: gb\_est3: \*  
23: gb\_est4: \*  
24: gb\_est5: \*  
25: gb\_est6: \*  
26: gb\_est7: \*  
27: gb\_est8: \*  
28: gb\_est9: \*  
29: gb\_est10: \*  
30: gb\_est11: \*  
31: gb\_est12: \*  
32: gb\_est13: \*  
33: gb\_est14: \*  
34: gb\_est15: \*  
35: gb\_est16: \*  
36: gb\_est17: \*  
37: gb\_est18: \*  
38: gb\_est19: \*  
39: gb\_est20: \*  
40: gb\_est21: \*  
41: gb\_est22: \*  
42: gb\_est23: \*  
43: gb\_est24: \*  
44: gb\_est25: \*

45: gb\_est26: \*  
46: gb\_est27: \*  
47: gb\_est28: \*  
48: gb\_est29: \*  
49: gb\_est30: \*  
50: gb\_est31: \*  
51: gb\_est32: \*  
52: em\_est20: \*  
53: em\_est21: \*  
54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*  
59: gb\_est33: \*  
60: gb\_est34: \*  
61: gb\_est35: \*  
62: gb\_est36: \*  
63: gb\_est37: \*  
64: gb\_est38: \*  
65: em\_est27: \*  
66: em\_est28: \*  
67: em\_est29: \*  
68: em\_est30: \*  
69: gb\_est39: \*  
70: gb\_est40: \*  
71: gb\_est41: \*  
72: gb\_est42: \*  
73: gb\_est43: \*  
74: gb\_est44: \*  
75: em\_est31: \*  
76: em\_est32: \*  
77: em\_est33: \*  
78: em\_est34: \*  
79: gb\_est45: \*  
80: gb\_est46: \*  
81: gb\_est47: \*  
82: gb\_est48: \*  
83: gb\_est49: \*  
84: gb\_est50: \*  
85: gb\_est51: \*  
86: em\_est35: \*  
87: em\_est36: \*  
88: em\_est37: \*  
89: em\_est38: \*  
90: gb\_est52: \*  
91: gb\_est53: \*  
92: gb\_est54: \*  
93: gb\_est55: \*  
94: gb\_est56: \*  
95: em\_est39: \*  
96: em\_est40: \*  
97: em\_est41: \*  
98: em\_est42: \*  
99: em\_est43: \*  
100: em\_est44: \*  
101: em\_est45: \*  
102: gb\_est57: \*  
103: gb\_est58: \*  
104: em\_est46: \*  
105: gb\_est59: \*  
106: gb\_est60: \*  
107: gb\_est61: \*  
108: gb\_est62: \*  
109: gb\_est63: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result

\*  
Query

SUMMARIES

No.	Score	Match	Length	DB	ID	Description
1	17.8	44.5	64	39	AA904844	AA904844 oJ80f12.s
2	17.6	44.0	56	70	AM156609	AM156609 se27e11.y
3	17	42.5	95	20	T52775	T52775 ya79g08.r1
4	16.6	41.5	73	37	AA683533	AA683533 zf34c12.s
5	16.2	40.5	55	42	AI079073	AI079073 SWBMLJ5BH
6	16	40.0	75	28	AA103683	AA103683 mo40b04.r
7	16	40.0	88	44	AI125675	AI125675 u130a03.y
8	16	40.0	90	37	AA688350	AA688350 nv15a05.s
9	15.8	39.5	79	21	T87474	T87474 yq85e02.r1
10	15.6	39.0	60	44	AI138205	AI138205 ta52c01.x
11	15.6	39.0	99	41	AU011571	AU011571 AU011571
12	15.4	38.5	100	22	R58627	R58627 G4408 fetal
13	15.2	38.0	67	39	AA866073	AA866073 oh31h05.s
14	15.2	38.0	90	41	AU007806	AU007806 AU007806
15	15.2	38.0	90	41	AU007807	AU007807 AU007807
16	15.2	38.0	90	41	AF083179	AF083179 Capra_hir
17	15	37.5	88	39	AA864661	AA864661 oh37c11.s
18	15	37.5	90	34	AA503993	AA503993 nh39e01.s
19	15	37.5	94	43	AI172856	AI172856 uc10c09.r
20	15	37.5	94	62	AI936176	AI936176 wo62q09.x
21	15	37.5	97	41	AI034469	AI034469 ow24d12.x
22	15	37.5	98	23	H39547	H39547 y168h07.r1
23	14.8	37.0	65	38	AA755665	AA755665 vs74e05.r
24	14.8	37.0	77	34	AA494110	AA494110 ne54b01.s
25	14.8	37.0	77	39	AA839381	AA839381 vv48h02.r
26	14.6	36.5	49	33	AA433110	AA433110 vq94g03.r
27	14.6	36.5	67	47	AI491935	AI491935 to06g01.x
28	14.6	36.5	69	21	R06240	R06240 y694h09.r1
29	14.6	36.5	74	45	AJ237129	AJ237129 AJ237129
30	14.6	36.5	100	40	AA928885	AA928885 v244d11.r
31	14.4	36.0	52	20	D25845	D25845 HUMGS04217
32	14.4	36.0	58	27	C00161	C00161 HUMGS000581
33	14.4	36.0	61	28	AA091692	AA091692 m1168.seq
34	14.4	36.0	64	37	AA716059	AA716059 zh32f12.s
35	14.4	36.0	73	39	AA878250	AA878250 o104f02.s
36	14.4	36.0	76	25	N89340	N89340 K9244F.Huma
37	14.4	36.0	79	44	AI272143	AI272143 ap61e04.x
38	14.4	36.0	81	23	H42950	H42950 y068c07.r1
39	14.4	36.0	81	23	R86045	R86045 y062a01.r1
40	14.4	36.0	82	36	AA664777	AA664777 nu68f02.s
41	14.4	36.0	82	62	AI906839	AI906839 RC-BT126-
42	14.4	36.0	82	62	AI924362	AI924362 wn55h10.x
43	14.4	36.0	88	62	AI902577	AI902577 QV-BT009-
44	14.4	36.0	91	21	T81025	T81025 yd25d10.s1
45	14.4	36.0	91	21	T81025	T81025 yd25d10.s1

## ALIGNMENTS

RESULT	1	ALIGNMENTS
AA904844	64 bp	09-JUN-1998
LOCUS	oJ80f12.s1	Seares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION	IMAGE:1504655 3' similar to TR:000405 000405 FB19 PROTEIN. ;, mRNA sequence.	
ACCESSION	AA904844	
VERSION	AA904844.1	GI:3039967
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
	Eutheria; Primates; Catarrhini; Homiidae; Homo.	
REFERENCE	1 (bases 1 to 64)	
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),	
JOURNAL	Tumor Gene Index	
COMMENT	Unpublished (1997)	
	Contact: Robert Strausberg, Ph.D.	
	Tel: (301) 496-1550	
	Email: Robert_Strausberg@nih.gov	

This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
 Trace considered overall poor quality  
 Insert length: 383 Std Error: 0.00  
 Seq primer: -40m13 fwd ER from Amersham  
 High quality sequence stop: 1.  
 Location/Qualifiers  
 1. 64

## FEATURES

source  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1504655"  
 /clone\_lib="Seares\_NFL\_T\_GBC\_S1"  
 /lab\_host="DH10B"  
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site:1: Not I; Site:2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NHH19W, testis NHR, and B-cell NCI-CGAP-GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."

## BASE COUNT

9 a 14 c 31 g 10 t

Query Match 44.5%; Score 17.8; DB 39; Length 64;  
 Best Local Similarity 67.6%; Pred. No. 6.1e-02;  
 Matches 25; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Oy 4 ggcccgccgtgtcatcattggcagcggtgaatt 40  
 Db 10 gccctggccgggaatttggcccccggggaattt 46

RESULT	2	
LOCUS	AM156609	56 bp
DEFINITION	se27e11.y1 Gm-c1015 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:	
ACCESSION	AM156609	
VERSION	AM156609.1	GI:6227947
KEYWORDS	EST.	
SOURCE	soybean.	
ORGANISM	Glycine max	
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core	
	eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;	
	Glycine.	

## REFERENCE

1 (bases 1 to 56)  
 Shoemaker R., Keim P., Vodkin L., Eipelting J., Coryell V.,  
 Khanna A., Bolla B., Marra M., Hillier L., Kucaba T., Martin J.,  
 Beck C., Wylie T., Underwood K., Steptoe M., Thelshing B., Allen M.,  
 Bowers Y., Person B., Swaller T., Gibbons M., Pape D., Harvey N.,  
 Schurk R., Ritter E., Kohn S., Shih T., Jackson Y., Cardenas M.,  
 McCann R., Materston R. and Wilson R.

Public Soybean EST Project  
 Unpublished (1999)  
 On Jul 8, 1999 this sequence version replaced gi:5421748.  
 Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project

Washington University School of Medicine  
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: est@watson.wustl.edu  
 This clone is available through: Genome Systems, Inc. 4633 World  
 Parkway Circle St. Louis, Missouri 63134 For further information  
 call: (800) 430-0030 or (314) 427-3322 FAX: (888) 919-3324 or (314)  
 427-3324 or contact: clones@genomesys.com or



TGTACCAATGTGAAGTGGAGCGCCGACATCTTTTCTTTTCTTTT 3',  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT73 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by  
 M.Felina Bonaldo. This library was constructed from the  
 same fetus as the fetal lung library, Soares fetal lung  
 NBHL19W."

BASE COUNT 18 a 17 c 13 g 25 t  
 ORIGIN

Query Match 41.5%; Score 16.6; DB 37; Length 73;  
 Best Local Similarity 71.0%; Pred. No. 2.1e+03;  
 Matches 22; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 10 ggcgtgataatgcatgacgagtaagt 40  
 ||| ||| | ||||| | ||| |||||  
 Db 73 GGCAGGTGACCATGTGCACGCTAGAACTT 43

RESULT 5  
 LOCUS A1079073 56 bp mRNA EST 12-AUG-1998  
 DEFINITION SWBML3BHL1G10SK Brugia malayi L3 subtracted cDNA library  
 (SAM97YG-Bml3SB) Brugia malayi cDNA clone SWBML3BHL1G10 5' or 3',  
 mRNA sequence.  
 ACCESSION A1079073  
 VERSION A1079073.1 GI:3415324  
 KEYWORDS EST.  
 SOURCE Brugia malayi.  
 ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;  
 Brugia malayi.  
 Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spirurida;  
 Filarioidea; Onchocercidae; Brugia.  
 REFERENCE 1 (bases 1 to 56)  
 AUTHORS Williams, S.A.  
 TITLE Genes expressed in infective third stage larvae of Brugia malayi  
 JOURNAL Unpublished (1995)  
 COMMENT On May 8, 1995 this sequence version replaced gi:801111.  
 CONTACT: Steven A. Williams  
 MOLECULAR PARASITOLOGY  
 Smith College Department of Biological Sciences  
 Department of Biological Sciences, Clark Science Center, Smith  
 College, Northampton, MA, 01063, USA  
 Tel: 4135853826  
 Fax: 4135853786  
 Email: genome@smith.edu  
 Seq primer: bluescript SK.  
 FEATURES  
 source Location/Qualifiers  
 1..56  
 /organism="Brugia malayi"  
 /db\_xref="taxon:6279"  
 /clone="SWBML3BHL1G10"  
 /clone\_1lb="Brugia malayi L3 subtracted cDNA library  
 (SAM97YG-Bml3SB)"  
 /dev\_stage="conventional L3 cDNA minus spliced-leader L3  
 cDNA"  
 /lab\_host="XL1-Blue MRF"  
 /note="Vector: Lambda ZAP II (stratagene); Site\_1: Eco RI;  
 Site\_2: Eco RI; Lymphatic filarial nematode parasite of  
 humans. cDNA inserts were obtained by subtracting the  
 JH093SL-Bml3 cDNA library (constructed using SL sequence  
 and PCR) from SAM94WL-Bml3 cDNA library (constructed  
 using conventional cDNA methods). cDNA was amplified from  
 both libraries using PCR and SL cDNA clones were  
 eliminated through hybridization. The library has 5 x  
 10E6 independent recombinants and the average insert size  
 is ~ 600 bp. The library was constructed by Ying Gong.  
 The library is available from Dr.S.A.Williams, email:  
 genome@smith.edu."

BASE COUNT 16 a 12 c 11 g 17 t  
 ORIGIN

Query Match 40.5%; Score 16.2; DB 42; Length 56;  
 Best Local Similarity 72.4%; Pred. No. 2.9e+03;  
 Matches 21; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 gacggccgggctgtatcaatggcata 29  
 | | ||||| | | | | | | | | | | |  
 Db 7 GGCAGGTGACCATGTGCACGCTAGAACTT 35

RESULT 6  
 LOCUS AAI03683 75 bp mRNA EST 29-OCT-1996  
 DEFINITION m040b04.r1 life Tech mouse embryo 15 5dpc 10667012 Mus musculus  
 cDNA clone IMAGE:556015 5', mRNA sequence.  
 ACCESSION AAI03683  
 VERSION AAI03683.1 GI:1649846  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 75)  
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.  
 TITLE The WashU-HMNI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT On Nov 29, 1993 this sequence version replaced gi:430148.  
 CONTACT: Marra M/Mouse EST Project  
 WashU-HMNI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LBNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:336807  
 Seq primer: -28M13 rev1 from Amersham  
 High quality sequence stop: 72.  
 FEATURES  
 source Location/Qualifiers  
 1..75  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:556015"  
 /clone\_1lb="Life Tech mouse embryo 15 5dpc 10667012"  
 /tissue\_type="embryo"  
 /dev\_stage="15.5dpc embryos"  
 /lab\_host="DH10B"  
 /note="Organ: whole embryo; Vector: PCMV-SPORT2; Site\_1:  
 SalI; Site\_2: NotI; Cloned unidirectionally. Primer:  
 Oligo dT. 15.5dpc embryos. PCMV-SPORT2 vector."  
 BASE COUNT 20 a 16 c 19 g 20 t  
 ORIGIN

Query Match 40.0%; Score 16; DB 28; Length 75;  
 Best Local Similarity 79.2%; Pred. No. 3.8e+03;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 9 ggcgtgataatgcatgacgagc 32  
 ||||| ||||| | | | | | | | | | | |  
 Db 17 GGCAGGTGACCATGTGCACGCTAGAACTT 40

RESULT 7  
 LOCUS A1256675/c 88 bp mRNA EST 12-NOV-1998

DEFINITION	U130A03.y1 Soares mouse urogenital ridge NMUR Mus musculus CDNA clone IMAGE:1852780 5', mRNA sequence.
ACCESSION	A1256675
VERSION	A1256675.1 GI:3864200
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le'M., Martin,J., Morris,M., Schellenberg,K., Steptoe,W., Tan,F., Underwood,K., Moore,B., Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Warston,R.
TITLE	The MashU-HMM Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On Jan 19, 1998 this sequence version replaced gi:12284957. Contact: Marra M/Mouse EST Project WashU-HMM Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: mouses@wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium ( <a href="#">info@image.llnl.gov</a> ) for further information. MGI:968208 Seq primer: -40RP from Gibco High quality sequence stop: 58. Location/Organism: . .88 /organism="Mus musculus" /db_xref="taxon:10090" /clone="IMAGE:1852780" /clone_lib="Soares mouse urogenital ridge NMUR" /sex="equal ratio of male:female" /tissue_type="urogenital ridge (embryonic)" /dev_stage="fetal, mixture of 11.5 and 12.5 dpc" /lab_host="DH10B" /note="Organ: gonad; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTCACATCTGAGTGCGAGCCGCAGTCATTCTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia). digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT	21 a 25 c 18 g 24 t
ORIGIN	
Query Match	40.0%; Score 16; DB 44; Length 88;
Best Local Similarity	62.5%; Pred. NO. 4e+03;
Matches	25; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Cy	1 gagggccggtgtatcatcattgcagtagcgagtctt 40   Db 77 GCCTGTCCGAGGGTGTGCCCTTGAATCCTGCTGACT 38
RESULT 8	
LOCUS	AA688350 90 bp mRNA EST 11-DEC-1997
DEFINITION	nvl5a05.s1 NC1-GAP_P122 Homo sapiens CDNA clone IMAGE:1220240 3', mRNA sequence.
ACCESSION	AA688350
VERSION	AA688350.1 GI:2675256
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```

REFERENCE      Eutheria: Primates: Catarrhini; Hominiidae; Homo.
AUTHORS       1 (bases 1 to 90)
TITLE         NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
              National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
JOURNAL       Unpublished (1997)
COMMENT       On Nov 6, 1997 this sequence version replaced gi:315230.
              Contact: Robert Strausberg, Ph.D.
              Tel: (301) 496-1550
              Email: Robert_Strausberg@nih.gov
              Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
              Emmert-Buck, M.D., Ph.D.
              CDNA Library Preparation: M. Bento Soares, Ph.D.
              CDNA Library Arrayed by: Greg Lennon, Ph.D.
              DNA sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              www-bio.llnl.gov/bdrip/image/image.html

FEATURES
Source        Seq primer: -40m13 fwd. ET from Amersham
              High quality sequence stop: 78.
              Location/Qualifiers
                1..90
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:1220240"
                  /clone_1lb="NCI_CGAP_Pr22"
                  /sex="male"
                  /tissue_type="normal prostate"
                  /lab_host="DH10B"
                  /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
                  with a modified polylinker; 1st strand cDNA was prepared
                  from normal prostate bulk tissue, and was then primed with
                  a Not I - oligo(dT) primer. Double-stranded cDNA was
                  ligated to Eco RI adaptors (Pharmacia), digested with Not
                  I and cloned into the Not I and Eco RI sites of the
                  modified pT73 vector. Library is normalized, and was
                  constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT    36 a          6 c          17 g          31 t

ORIGIN
Query Match   40.0%; Score 16; DB 37; Length 90;
Best Local Similarity 79.2%; Pred. No. 4e+03;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

CY 14 ggtatcaattggcagtcagcggtaa 37
      1 ||||| 11111111111111111111
DB 48 gctatcaattggamcagcgcgca 71

RESULT 9
LOCUS      T87474
DEFINITION yd85e02.r1 Soares fetal liver spleen INFLS Homo sapiens CDNA clone
VERSION    IMAGE:115034 5' similar to SP:S27875 S27875 GENE PC326 PROTEIN - ;
KEYWORDS   mRNA sequence.
ACCESSION  T87474
VERSION    T87474.1 GI:715826
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 79)
AUTHORS   Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
            Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
            Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
            Trevisakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
            Wilson, R.
            The WashU-Merck EST Project
            Unpublished (1995)
            Contact: Wilson RK

```

found through the I.M.A.G.E. Consortium  
www.bio.llnl.gov/bbrp/image/image.html

BASE COUNT	the World Wide Web.	(URL, http://www.nirs.go.jp)"
27 a	31 c	16 g
		24 t
		1 others

BASE COUNT	the World Wide Web.	(URL, http://www.nirs.go.jp)"
27 a	31 c	16 g
		24 t
		1 others





**FEATURES**  
source  
National Institute of Radiological Sciences  
9-1, Anagawa-4-chome, Inage-ku, Chiba 263-8555, Japan  
Email: mori@nri.90.jp.  
Location/Qualifiers  
1..90

```

source
1. .90
/organism="Schizosaccharomyces pombe"
/strain="972"
/db_xref="taxon:4896"
/clone="spc02527"
/clone_1fp="Schizosaccharomyces pombe late log phase cDNA"
/sex="h minus"
/note="vector: M13mp19. The cDNA library of
Schizosaccharomyces pombe was prepared by cloning cDNA
into the SmaI site of M13mp19 DNA and the direction of DNA
sequences was not always from 5' to 3'. The cDNA data of
Schizosaccharomyces pombe are available for searching on
the World Wide Web. (URL: http://www.nllrs.go.jp)"

```

38	Query Match	38.0%	Score 15.2	DB 41	Length 90
39	Best Local Similarity	71.4%	Pred. No. 8.9e+03		
40	Matches 20	Conservative	0	Mismatches 8	Indels 0
41	QY	10	gacgtgtatcaatcgtgcataaggcgttaa	37	
42					
43					
44		68	GCCTGTGTTCTGCTGCAACGGCGGAAA	41	

RESULT	15
AU007807/c	
LOCUS	AU007807 90 bp mRNA EST 31-Jul-1998
DEFINITION	AU007807 Schizosaccharomyces pombe late log phase cDNA
ACCESSION	Schizosaccharomyces pombe cDNA clone spc02528, mRNA sequence.
VERSION	AU007807
KEYWORDS	AU007807.1 GI:3344265
SOURCE	EST.
ORGANISM	fission yeast. Schizosaccharomyces pombe

REFERENCE	1 (bases 1 to 90)
AUTHORS	Moriygo,M. and Mita,K.
TITLE	Identification of expressed sequence tags of Schizosaccharomyces pombe
JOURNAL	Unpublished (1998)
COMMENT	On May 9, 1995 this sequence version replaced gi:803109.

**JOURNAL** Unpublished (1998)  
**COMMENT** On May 9, 1995 this sequence version replaced gi:803109.  
Contact: Mitsuoki Moriyomo  
Genome Research Group  
National Institute of Radiological Sciences  
9-1, Anagawa-4-chome, Inage-ku, Chiba 263-8555, Japan  
Email: moriyomo@nirs.go.jp.  
Location/Qualifiers  
1. .90

```

/clone_11p="Schizosaccharomyces pombe late log phase cDNA"
/sex="h minus"
/note="Vector: M13mp19. The cDNA library of
Schizosaccharomyces pombe was prepared by cloning cDNA
into the SmaI site of M13mp19 DNA and the direction of DNA
sequences was not always from 5' to 3'. The cDNA data of
Schizosaccharomyces pombe are available for searching on
the World Wide Web. (URL, http://www.nlrs.go.jp)"
BASE COUNT
ORIGIN
24 a      29 c      16 g      21 t

```

Query Match	38.0%;	Score 15.2;	DB 41;	Length 90;
Best Local Similarity	71.4%;	Pred. No. 8.9e+03;		

	Matches	20; Conservative	0; Mismatches	8; Indels	0; Gaps
QY	10	ggctggtatcaattgcatagcggttaa	37		
Db	68	ggcttgtttcttcgttcgcaacggcgcaaa	41		

Search completed: June 4, 2000, 13:53:32  
Job time: 20956 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:09:46 ; Search time 244.64 Seconds  
(without alignments)  
21.253 Million cell updates/sec

Title: US-09-164-714-17

Perfect score: 40  
Sequence: 1 gacggccggcgctgtatcaattgcataaggcgtaagt 40

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64392525 residues

Total number of hits satisfying chosen parameters: 375880

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/6.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/ina/6.COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	15.8	39.5	38	5	US-08-718-904-124
2	15.6	39.0	51	3	US-08-432-871C-25
3	15	37.5	94	2	US-08-472-255A-24
4	15	37.5	94	2	US-08-479-724A-24
5	15	37.5	94	5	US-08-472-256B-24
6	15	37.5	94	6	PCT-US96-09455A-24
7	14.6	36.5	35	2	US-08-446-856A-2
8	14.4	36.0	93	2	US-08-388-672A-11
9	14.4	36.0	93	5	US-09-080-554-11
10	14.4	36.0	97	6	PCT-US96-09455A-284
11	14.2	35.5	49	1	US-08-298-073-22
12	14.2	35.5	49	2	US-08-794-153-22
13	14.2	35.5	73	2	US-08-434-001-64
14	14.2	35.5	73	2	US-08-433-585-64
15	14.2	35.5	73	2	US-08-434-425-64
16	14.2	35.5	73	2	US-08-437-667-64
17	14.2	35.5	73	5	US-08-906-955-64
18	14.2	35.5	73	6	PCT-US96-06060-64
19	14	35.0	22	2	US-08-468-580-20
20	14	35.0	22	6	PCT-US95-03731-20
21	14	35.0	31	1	US-08-457-245-18
22	14	35.0	37	3	US-08-880-557-35
23	14	35.0	63	1	US-08-343-943-13
24	14	35.0	96	6	PCT-US96-09455A-274
25	13.8	34.5	23	1	US-08-010-099-6
26	13.8	34.5	23	1	US-08-010-099-15
27	13.8	34.5	23	1	US-08-010-099-15

c 28	13.8	34.5	23	1	US-08-010-099-22	Sequence 22, App1
c 29	13.8	34.5	23	2	US-08-448-716-6	Sequence 6, App1
c 30	13.8	34.5	23	2	US-08-448-716-15	Sequence 15, App1
c 31	13.8	34.5	23	2	US-08-448-716-18	Sequence 18, App1
c 32	13.8	34.5	23	2	US-08-448-716-22	Sequence 22, App1
c 33	13.8	34.5	55	6	PCT-US96-09451-175	Sequence 71, App1
c 34	13.6	34.5	77	1	US-08-399-412A-14	Sequence 14, App1
c 35	13.6	34.0	48	1	US-08-140-797-2	Sequence 2, App1
c 36	13.6	34.0	48	1	US-08-486-670A-2	Sequence 2, App1
c 37	13.6	34.0	66	2	US-08-323-531-34	Sequence 34, App1
c 38	13.6	34.0	66	2	US-08-198-094-34	Sequence 34, App1
c 39	13.6	34.0	66	6	PCT-US93-07424-34	Sequence 34, App1
c 40	13.6	34.0	66	6	PCT-US95-02087-34	Sequence 34, App1
c 41	13.6	34.0	78	1	US-08-411-796-175	Sequence 175, App1
c 42	13.6	34.0	78	5	US-08-471-039-175	Sequence 175, App1
c 43	13.6	34.0	78	6	PCT-US93-11198-175	Sequence 175, App1
c 44	13.4	33.5	43	2	US-08-292-081A-13	Sequence 13, App1
c 45	13.4	33.5	43	3	US-08-752-495-13	Sequence 13, App1

## ALIGNMENTS

RESULT 1

US-08-718-904-124  
Sequence 124, Application US/08718904  
Patent No. 6037329

GENERAL INFORMATION:

APPLICANT: Baird, J. Andrew

APPLICANT: Chandler, Lois Ann

APPLICANT: Sosnowski, Barbara A.

TITLE OF INVENTION: COMPOSITIONS CONTAINING NUCLEIC ACIDS AND LIGANDS FOR THERA

NUMBER OF SEQUENCES: 128

CORRESPONDENCE ADDRESSES:

ADDRESS: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/718,904

FILING DATE: 24-SEP-1996

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: NO. 6037329tenburg Ph.D., Carol

REGISTRATION NUMBER: 39,317

REFERENCE/DOCKET NUMBER: 760100.415C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 124:

SEQUENCE CHARACTERISTICS:

LENGTH: 38 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-718-904-124

Query Match 39.5%; Score 15.8; DB 5; Length 38;  
Best Local Similarity 74.1%; Pred. No. 98;  
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
DB 8 TGTCTCATTCCTTGACCGTAAGT 34

```

RESULT 2
US-08-432-871C-25
Sequence 25, Application US/08432871C
Patent No. 5877010
GENERAL INFORMATION:
APPLICANT: Loeb, Lawrence A.
APPLICANT: Black, Margaret E.
TITLE OF INVENTION: THYMIDINE KINASE MUTANTS
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESSES:
ADDRESS: Seed and Berry LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/432,871C
FILING DATE: 02-MAY-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: McMasters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 240052.409C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-432-871C-25

Query Match 39.0%; Score 15.6; DB 3; Length 51;
Best Local Similarity 53.8%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 18; Indels 0; Gaps 0.

QY 2 acggccgggctggtatcaattgcatggcggttaagt 40
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8 ACCGGCGGCGCGGGTACGANNNNNNNNNGGCGATGGGAT 46

RESULT 3
US-08-472-255A-24
Sequence 24, Application US/08472255A
Patent No. 5766853
GENERAL INFORMATION:
APPLICANT: PARMA, DAVID
APPLICANT: GOLD, LARRY
TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS
TITLE OF INVENTION: TO SELECTINS (AS AMENDED)
NUMBER OF SEQUENCES: 173
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MB
COMPUTER: IBM pc compatible
OPERATING SYSTEM: MS-DOS

```

```

1 SOFTWARE: WordPerfect 6.0
2 CURRENT APPLICATION DATA:
3 APPLICATION NUMBER: US/08/472,255A
4 FILING DATE: 07-JUNE-1995
5 CLASSIFICATION: 536
6 PRIOR APPLICATION DATA:
7 APPLICATION NUMBER: 07/714,131
8 FILING DATE: 10-JUNE-1991
9 PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: 07/536,428
11 FILING DATE: 11-JUNE-1991
12 PRIOR APPLICATION DATA:
13 APPLICATION NUMBER: 07/964,624
14 FILING DATE: 21-OCTOBER-1992
15 ATTORNEY/AGENT INFORMATION:
16 NAME: Barry J. Swanson
17 REGISTRATION NUMBER: 33,215
18 REFERENCE/DOCKET NUMBER: NEX40-3
19 TELECOMMUNICATION INFORMATION:
20 TELEPHONE: (303) 793-3333
21 TELEFAX: (303) 793-3433
22 INFORMATION FOR SEQ ID NO: 24:
23 SEQUENCE CHARACTERISTICS:
24 LENGTH: 94 base pairs
25 TYPE: nucleic acid
26 STRANDEDNESS: single
27 TOPOLOGY: linear
28 MOLECULE TYPE: RNA
29 FEATURE:
30 OTHER INFORMATION: All C's are 2'-NH2 cytosine
31 FEATURE:
32 OTHER INFORMATION: All U's are 2'-NH2 uracil
33
34 US-08-472-255A-24

```

Query Match	37.5%	Score 15;	DB 2;	Length 94;
Best Local Similarity	56.4%;	Pred. No. 2.7e+02;		
Matches 22;	Conservative 2;	Mismatches 15;	Indels 0;	Gaps 0;

```
QY      2  acgcccggcgtgtgataatcattggcatagcgcgtaagt 40
          |  ||||| ||| |  : ||| | || | | :
Db      35  AUGGCCGGGGCAGGAAGUGGUGGCCAACGAGGCCAGCU 73
```

RESULT  
 US-08-479-724A-24  
 Sequence 24, Application US/08479724A  
 Patent No. 5780228  
 GENERAL INFORMATION:  
 APPLICANT: PARMA, DAVID  
 APPLICANT: GOLD, LARRY  
 TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS  
 TITLE OF INVENTION: TO LECTRINS  
 NUMBER OF SEQUENCES: 173  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Swanson & Bratschun, L.L.C.  
 STREET: 8400 E. Prentice Avenue, Suite 200  
 CITY: Englewood  
 STATE: Colorado  
 COUNTRY: USA  
 ZIP: 80111  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MB  
 COMPUTER: IBM pc compatible  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: Wordperfect 6.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/479,724A  
 FILING DATE: 07-JUNE-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/714,131  
 FILING DATE: 10-JUNE-1991

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/536,428
; FILING DATE: 11-JUNE-1991
; PRIOR APPLICATION DATA: 07/964,624
; APPLICATION NUMBER: 07/964,624
; FILING DATE: 21-OCTOBER-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER: NEX40-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ. ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 94 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: RNA
; FEATURE:
; OTHER INFORMATION: All C's are 2'-NH2 cytosine
; FEATURE:
; OTHER INFORMATION: All U's are 2'-NH2 uracil
US-08-479-724A-24

Query Match          37.5%; Score 15; DB 2; Length 94;
Best Local Similarity 56.4%; Pred. No. 2.7e+02;
Matches 22; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 2 acggccgggctgtatcaattggcatagcggttaagt 40
      ||||| ||||| ||||| ||||| ||||| |||||
DB 35 AUGGCCGGGCGAGGAGGUGGUGGCAACGAGGCCAGCU 73

RESULT 5
US-08-472-256B-24
; Sequence 24, Application US/08472256B
; Patent No. 6001988
; GENERAL INFORMATION:
; APPLICANT: PARMA, DAVID
; APPLICANT: GOLD, LARRY
; TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS
; TITLE OF INVENTION: TO LECTINS
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson & Bratschun, L.L.C.
; STREET: 8400 E. Prentice Avenue, Suite 200
; CITY: Englewood
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MB
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,256B
; FILING DATE: 07-JUNE-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/714,131
; FILING DATE: 10-JUNE-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/536,428
; FILING DATE: 11-JUNE-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/964,624
; FILING DATE: 21-OCTOBER-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson

```

```

; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ. ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 94 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: RNA
; FEATURE:
; OTHER INFORMATION: All C's are 2'-NH2 cytosine
; FEATURE:
; OTHER INFORMATION: All U's are 2'-NH2 uracil
US-08-472-256B-24

Query Match          37.5%; Score 15; DB 5; Length 94;
Best Local Similarity 56.4%; Pred. No. 2.7e+02;
Matches 22; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 2 acggccgggctgtatcaattggcatagcggttaagt 40
      ||||| ||||| ||||| ||||| ||||| |||||
DB 35 AUGGCCGGGCGAGGAGGUGGUGGCAACGAGGCCAGCU 73

RESULT 6
PCT-US96-09455A-24
; Sequence 24, Application PC/TUS9609455A
; GENERAL INFORMATION:
; APPLICANT: PARMA, et al.
; TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID
; TITLE OF INVENTION: LIGANDS TO LECTINS
; NUMBER OF SEQUENCES: 390
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson & Bratschun, L.L.C.
; STREET: 8400 E. Prentice Avenue, Suite 200
; CITY: Englewood
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MB
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/09455A
; FILING DATE: 05 JUNE 1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/479,724
; FILING DATE: 07-JUNE-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/472,256
; FILING DATE: 07-JUNE-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/472,255
; FILING DATE: 07-JUNE-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/477,829
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER: NEX40C/PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ. ID NO: 24:
; SEQUENCE CHARACTERISTICS:

```

```

?      LENGTH:   94 base pairs
?      TYPE:     nucleic acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      MOLECULE TYPE: RNA
?      FEATURE:
? OTHER INFORMATION: ALL C's are 2'-NH2 cytosine
?      FEATURE:
? OTHER INFORMATION: ALL U's are 2'-NH2 uracil
?      OTHER INFORMATION: ALL U's are 2'-NH2 uracil
PCT-US96-09455A-24

```

Query Match	37.5%	Score 15:	DB 6:	Length 94:
Best Local Similarity	56.4%	Pred. No.	2.7e+02:	
Matches 22; Conservative		Mismatches 15;	Indels 0;	Gaps 0

QY 2 acgagcccgagctgtatcaattgycatagcgtaagt 40  
| ||| ||| || :||| || || :  
Db 35 AUGGCGGCGGCAGAGUGGUGGCCAACGAGGCCAGCU 73

RESULT 7  
US-08-446-856A-2  
; Sequence 2, Application US/08446856A

GENERAL INFORMATION: Peter Michael  
APPLICANT: Broad,  
TITLE OF INVENTION: Identification of Compounds  
Modulating Protein/Cell Membrane Interaction  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pilsbury Madison & Sutoro, L.L.P.  
STREET: 1100 New York Avenue, N.W.

```

1 STATE: Washington
2
3 STATE: D.C.
4
5 COUNTRY: USA
6
7 ZIP: 20005-3918
8
9 COMPUTER READABLE FORM:
10
11 MEDIUM TYPE: Floppy disk
12
13 COMPUTER: IBM PC compatible
14
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16
17 SOFTWARE: Word Perfect
18
19 CURRENT APPLICATION DATA:
20
21 APPLICATION NUMBER: US/08/446,856A
22
23 FILING DATE: 05-JUN-1995
24
25 CLASSIFICATION: 435
26
27 PRIOR APPLICATION DATA:
28
29 APPLICATION NUMBER: PCT/GB93/02543
30
31 FILING DATE: 14-DEC-1993
32
33 PRIOR APPLICATION DATA:
34
35 APPLICATION NUMBER: GB 9226065.2
36
37 FILING DATE: 14-DEC-1992
38
39 INFORMATION FOR SEQ ID NO: 2:
40
41 SEQUENCE CHARACTERISTICS:
42
43 LENGTH: 35 base pairs
44
45 TYPE: nucleic acid
46
47 STRANDEDNESS: single
48
49 TOPOLOGY: linear
50
51 US-08-446-856A-2

```

Query Match	36.5%	Score 14.6	DB 2	Length 35
Best Local Similarity	81.0%	Pred. No. 3.3e+02		
Matches 17	Conservative 0	Mismatches 4	Indels 0	Gaps 0

```

QY      4  gcccggctgtatcatctg 24
          |||||  |||||  |||||  |||
Db      7  GGGCCAGCTGTGATCAATTG 27

```

RESULT 8  
US-08-388-672A-11  
; Sequence 11, Application US/08388672A

Patent NO. 5795961  
GENERAL INFORMATION:  
APPLICANT: Wallace, T. Paul  
APPLICANT: Harris, William J.  
APPLICANT: Carr, Frank J.  
APPLICANT: Old, Lloyd J.  
APPLICANT: Welf, Sydney  
APPLICANT: Kitamura, Kunio  
TITLE OF INVENTION: Recombinant Human Anti-Lewis B  
TITLE OF INVENTION: Antibodies  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Felfe and Lynch  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10022  
COMPUTER READABLE FORM:

```

: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC compatible
:
: OPERATING SYSTEM: PC-DOS/MS-DOS
:
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:

```

;; APPLICATION NUMBER: US/08/388,672A  
;; FILING DATE: 14-FEB-1995

CLASSIFICATION:  
;  
ATTORNEY/AGENT INFORMATION:  
;

```

; NAME: Hanson, No. 5795961man D.
; REGISTRATION NUMBER: 30,946

```

REFERENCE/DOCKET NUMBER: LUD 5405  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-688-9200  
TELEFAX: 212-838-3884

```
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
```

```

; LENGTH: 93 base pairs
; TYPE: nucleic acid

```

```

; STRANDEDNESS: unknown
; TOPOLOGY: unknown

```

```

; MOLECULE TYPE: DNA (genomic)
US-08-388-672A-11

```

Query Match	36.0%	Score 14.4	DB 2	Length 93
Best Local Similarity	65.6%	Pred. No. 5e+02		
Matches 21; Conservative	0	Mismatches 11	Indels 0	Gaps 0

Qy 4 ggccggcgtgtatcaattgcatagcgct 35  
||| ||||| ||| ||||| |  
Db 4 GGCTTCTGCTGGTACCAATATAATAGGT 35

RESULT 9  
US-09-080-554-11  
; Sequence 11, Application US/09080554

? GENERAL INFORMATION:  
 ? APPLICANT: Wallace, T. Paul  
 ? APPLICANT: Harris, William J.  
 ? APPLICANT: Carr, Frank J.  
 ? APPLICANT: Old, Lloyd J.  
 ? APPLICANT: Welt, Sydney  
 ? APPLICANT: Kikamuta, Kunio  
 ? TITLE OF INVENTION: Recombinant Human Anti-Lewis B  
 ? TITLE OF INVENTION: Antidotes  
 ? NUMBER OF SEQUENCES: 25  
 ? CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.



Query Match	36.0%	Score 14.4	DB 5	Length 93
Best Local Similarity	65.6%	Pred. No. 5e+02		
Matches	21	Conservative	0	Mismatches 11
				Indels 0
				Gaps 0
OY	4	ggcccgagctggtalcaattgcataagcggt	35	
db	4	ggcttctgctgtaccacatataaataaggcgtt	35	

```

PRTOR APPLICATION DATA:
APPLICATION NUMBER: 08/477, 829
FILING DATE: 07-JUNE-1995
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX40C/PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 284:
SEQUENCE CHARACTERISTICS:
LENGTH: 97 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA
FEATURE:
OTHER INFORMATION: All C's are 2'-NH2 cytosine
PCT-US96-09455A-284

Query Match 36.0%; Score 14.4; DB 6; Length 97;
Best Local Similarity 56.2%; Pred. No. 5.1e+02;
Matches 18; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Oy 2 acggccgggctgcatgcatggatggcg 33
| | | | | | | | | | | | | | | | | | | | |
Db 22 AAGUACGGGAGGAGGAGUACUGGGAUAGG 53

RESULT 11
US-08-298-073-22
Sequence 22, Application US/08298073
Patent No. 5681/02
GENERAL INFORMATION:
APPLICANT: COLLINS, MARK L.
APPLICANT: HORN, THOMAS
APPLICANT: SHERIDAN, PATRICK E.
APPLICANT: WARNER, BRIAN D.
APPLICANT: URDEA, MICHAEL S.
TITLE OF INVENTION: REDUCTION OF NONSPECIFIC HYBRIDIZATION
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHIRON CORPORATION, INTELLECTUAL PROPERTY -
ADDRESS: R440
STREET: P.O. BOX 8097
CITY: EMERYVILLE
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94662-8097
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/298,073
FILING DATE: 30-AUG-1994
CLASSIFICATION: A35
ATTORNEY/AGENT INFORMATION:
NAME: GOLDMAN, KENNETH M.
REGISTRATION NUMBER: 34,174
REFERENCE/DOCKET NUMBER: 0974,001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2719
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 49 base pairs

```

```

;      TYPE: nucleic acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;      MOLECULE TYPE: DNA (genomic)
;
US-08-298-073-22

```

[illegible]

RESULT 12  
 US-08-794-153-22  
 Sequence 22, Application US/08/794153  
 Patent No. 5780610  
 GENERAL INFORMATION:  
 APPLICANT: COLLINS, MARK L.  
 APPLICANT: HORN, THOMAS  
 APPLICANT: SHERIDAN, PATRICK E.  
 APPLICANT: WARNER, BRIAN D.  
 APPLICANT: URDEA, MICHAEL S.  
 TITLE OF INVENTION: REDUCTION OF NONSPECIFIC HYBRIDIZATION  
 TITLE OF INVENTION: BY USING NOVEL, BASE-PAIRING SCHEMES  
 NUMBER OF SEQUENCES: 35  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: CATION CORPORATION, INTELLECTUAL PROPERTY -  
 ADDRESSEE: R440  
 STREET: P. O. BOX 8097  
 CITY: EMERYVILLE  
 STATE: CALIFORNIA  
 COUNTRY: UNITED STATES OF AMERICA  
 ZIP: 94662-8097  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/794,153  
 FILING DATE: 03-FEB-1997  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/298,073  
 FILING DATE: 30-AUG-1994  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: GOLDMAN, KENNETH M.  
 REGISTRATION NUMBER: 34,114  
 REFERENCE/DOCKET NUMBER: 0974,001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (510) 601-2719  
 TELEFAX: (510) 655-3542  
 INFORMATION FOR SEQ ID NO: 22:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 49 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-794-153-22

```
Query Match      35.5%   Score 14.2: DB 2; Length 49;
Best Local Similarity 84.2%; Pred. No. 5.4e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

Db 21 TTGTAATCATTAGGCATAGG 39

RESULT 13  
 US-08-434-001-64/c  
 Sequence 64, Application US/08434001  
 Patent No. 5712375  
 GENERAL INFORMATION:  
 APPLICANT: JENSEN, KIRK  
 APPLICANT: CHEN, HANG  
 APPLICANT: MORRIS, KEVIN  
 APPLICANT: STEPHENS, ANDREW  
 APPLICANT: GOLD, LARRY  
 TITLE OF INVENTION: SYSTEMATIC EVOLUTION OF LIGANDS BY  
 TITLE OF INVENTION: EXPERIMENTAL ENRICHMENT: TISSUE  
 TITLE OF INVENTION: SELEX  
 NUMBER OF SEQUENCES: 235  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Swanson & Bratschun, L.L.C.,  
 STREET: 8400 E. Prentice Avenue, Suite 200  
 CITY: Englewood  
 STATE: Colorado  
 COUNTRY: USA  
 ZIP: 80111  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MB  
 COMPUTER: IBM pc compatible  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: Wordperfect 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/434,001

```

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/714,131
FILING DATE: 10-JUNE-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/536,428
FILING DATE: 11-JUNE-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/964,624
FILING DATE: 21-OCTOBER-1992
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX30.3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 73 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

```

	Score	14.2;	DB	2;	Length	73;
Query Match:	62.9%;	Pred.	No.	5,	9e+02;	
Best Local Similarity	Conservative	0;	Mismatches	13;	Indels	0;
Matches	22;				Gaps	0;
QY	4	ggccgcgagctgtatcaattgcataggcggttaag	38			
		- - - - -				
b	67	GGCCGCATGTCGAATCACCGGAGTAGCGGGGAAG	33			
		- - - - -				

RESULT 14  
US-08-433-585-64/c  
; Sequence 64, Application US/08433585  
; Patent No. 5763566  
; GENERAL INFORMATION:  
; APPLICANT: JENSEN, KIRK

```

APPLICANT: CHEN, HANG
APPLICANT: MORRIS, KEVIN
APPLICANT: STEPHENS, ANDREW
APPLICANT: GOLD, LARRY
TITLE OF INVENTION: SYSTEMATIC EVOLUTION OF LIGANDS BY
TITLE OF INVENTION: EXPONENTIAL ENRICHMENT: TISSUE
TITLE OF INVENTION: SELEX
NUMBER OF SEQUENCES: 235
CORRESPONDENCE ADDRESSES:
ADDRESS: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MB
COMPUTER: IBM pc compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/433,585
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/714,131
FILING DATE: 10-JUNE-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/536,428
FILING DATE: 11-JUNE-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/964,624
FILING DATE: 21-OCTOBER-1992
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX30.4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 73 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-433-585-64

Query Match 35.5%; Score 14.2; DB 2; Length 73;
Best Local Similarity 62.9%; Pred. No. 5,9e+02;
Matches 22; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 ggcgcggcgtggtatcgaatggcgaatgagcggtgaag 38
||||| | | | | | | | | | | | | | | | | |
Db 67 ggccctcgtatgcgaatcaccgggatagccgggaag 33

RESULT 15
US-08-434-425-64/C
Sequence 64, Application US/08434425
Patent No. 5789157
GENERAL INFORMATION:
APPLICANT: JENSEN, KIRK
APPLICANT: CHEN, HANG
APPLICANT: MORRIS, KEVIN
APPLICANT: STEPHENS, ANDREW
APPLICANT: GOLD, LARRY
TITLE OF INVENTION: SYSTEMATIC EVOLUTION OF LIGANDS BY
TITLE OF INVENTION: EXPONENTIAL ENRICHMENT: TISSUE
TITLE OF INVENTION: SELEX
NUMBER OF SEQUENCES: 235
CORRESPONDENCE ADDRESSES:

```

```

1 ADDRESS: Swanson & Bratschun, L.L.C.
2 STREET: 8400 E. Prentice Avenue, Suite 200
3 CITY: Englewood
4 STATE: Colorado
5 COUNTRY: USA
6 ZIP: 80111
7
8 COMPUTER READABLE FORM:
9 MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MB
10 COMPUTER: IBM pc compatible
11 OPERATING SYSTEM: MS-DOS
12 SOFTWARE: Wordperfect 5.1
13
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER: US/08/434,425
16 FILING DATE:
17 CLASSIFICATION: 536
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: 07/714,131
20 FILING DATE: 10-JUNE-1991
21
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: 07/536,428
24 FILING DATE: 11-JUNE-1990
25
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: 07/964,624
28 FILING DATE: 21-OCTOBER-1992
29
30 ATTORNEY/AGENT INFORMATION:
31 NAME: Barry J. Swanson
32 REGISTRATION NUMBER: 33,215
33 REFERENCE/DOCKET NUMBER: NEX30.1
34 TELECOMMUNICATION INFORMATION:
35 TELEPHONE: (303) 793-3333
36 TELEFAX: (303) 793-3433
37
38 INFORMATION FOR SEQ ID NO: 64:
39 SEQUENCE CHARACTERISTICS:
40 LENGTH: 73 base pairs
41 TYPE: nucleic acid
42 STRANDEDNESS: single
43 TOPOLOGY: linear
44
45 US-08-434-425-64
46
47 Query Match 35.5%; Score 14.2; DB 2; Length 73;
48 Best Local Similarity 62.9%; Pred. No. 5.9e+02;
49 Matches 22; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
50
51 4 ggcgcgcgcgtatcatcattgcacagcgcgtcaag 38
52 ||||| | ||||| ||||| |||||
53 db 67 gcccttcattgcgaatcaccgcggatagccgggaag 33

```

Search completed: June 4, 2000, 16:09:47  
Job time: 28066 sec

```

QY      4  ggcccggtctgtatcatcaattgcatagcgcytaa 38
          |||  |  |||  |  |||  |  |||  |  |||
Db      67  GGCCTCATGTGCAATCACCCGGGATAGCCGGGAAG 33

```

RESULT 15  
US-08-434-425-64/C  
Sequence 64, Application US/08434425  
Patent No. 5789157  
GENERAL INFORMATION:  
APPLICANT: JENSEN, KIRK  
APPLICANT: CHEN, HANG  
APPLICANT: MORRIS, KEVIN  
APPLICANT: STEPHENS, ANDREW  
APPLICANT: GOLD, LARRY  
TITLE OF INVENTION: SYSTEMATIC EVOLUTION OF LIGANDS BY  
TITLE OF INVENTION: EXPONENTIAL ENRICHMENT: TISSUE  
TITLE OF INVENTION: SELEX  
NUMBER OF SEQUENCES: 235  
CORRESPONDENCE ADDRESS:

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:05:43 ; Search time 1236.38 Seconds  
(Without alignments)  
-29.899 Million cell updates/sec

Title: US-09-164-714-18

Perfect score: 38  
Sequence: 1 catgctgcagcttgaccacatttggcaccacaataggg 38

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 segs, -486395729 residues

Total number of hits satisfying chosen parameters: 356616

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pl4:\*  
11: gb\_pl5:\*  
12: gb\_pl6:\*  
13: gb\_pl7:\*  
14: gb\_pl8:\*  
15: gb\_pl9:\*  
16: gb\_pl10:\*  
17: gb\_pl11:\*  
18: gb\_pl12:\*  
19: gb\_pl13:\*  
20: gb\_pl14:\*  
21: gb\_pl15:\*  
22: gb\_pl16:\*  
23: gb\_pl17:\*  
24: gb\_pl18:\*  
25: gb\_pl19:\*  
26: gb\_pl20:\*  
27: gb\_pl21:\*  
28: gb\_pl22:\*  
29: gb\_pl23:\*  
30: gb\_pl24:\*  
31: gb\_pl25:\*  
32: gb\_pl26:\*  
33: gb\_pl27:\*  
34: gb\_pl28:\*  
35: gb\_pl29:\*  
36: gb\_pl30:\*  
37: gb\_pl31:\*  
38: gb\_pl32:\*  
39: gb\_pl33:\*  
40: gb\_pl34:\*  
41: gb\_pl35:\*  
42: gb\_pl36:\*  
43: gb\_pl37:\*  
44: gb\_pl38:\*

45: gb\_hlg7:\*  
46: em\_hlg1:\*  
47: em\_hlg2:\*  
48: em\_hlg3:\*  
49: em\_hlg4:\*  
50: gb\_pl13:\*  
51: gb\_pl14:\*  
52: gb\_pl15:\*  
53: gb\_pl16:\*  
54: gb\_pl17:\*  
55: gb\_pl18:\*  
56: gb\_pl19:\*  
57: gb\_pl20:\*  
58: gb\_pl21:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17.4	45.8	60	1	S77980	S77980 comp1-major
2	17.4	45.8	60	2	AF065436	AF065436 Chlamydia
3	17.4	45.8	60	2	CTB1AOMP1	AF070239 Chlamydia
4	17.4	45.8	60	2	CTB2AOMP1	AF070239 Chlamydia
5	17.4	45.8	60	2	CTB2AOMP1	AF070239 Chlamydia
6	17.4	45.8	60	2	CTB2AOMP1	AF070239 Chlamydia
7	17.4	45.8	60	2	CTB2AOMP1	AF070239 Chlamydia
8	17.4	45.8	60	2	CTB2AOMP1	AF070239 Chlamydia
9	17.4	45.8	60	2	CTB2AOMP1	AF070239 Chlamydia
10	17.4	45.8	60	2	CTB2AOMP1	AF070239 Chlamydia
11	17.4	45.8	60	2	CTB2AOMP1	AF070239 Chlamydia
12	17.4	45.8	60	2	CTB2AOMP1	AF070239 Chlamydia
13	17.4	45.8	60	2	CTB2AOMP1	AF070239 Chlamydia
14	17.4	45.8	60	2	CTB2AOMP1	AF070239 Chlamydia
15	17.4	45.8	60	2	CTB2AOMP1	AF070239 Chlamydia
16	17.4	45.8	60	2	CTB2AOMP1	AF070239 Chlamydia
17	17.4	45.8	60	2	CTB2AOMP1	AF070239 Chlamydia
18	17.4	45.8	60	2	CTB2AOMP1	AF070239 Chlamydia
19	17.4	45.8	100	12	MA286084	286084 M. auratus D
20	16.6	43.7	45	5	AR012219	AR012219 Sequence
21	16.6	43.7	45	5	I21230	I21230 Sequence 4
22	16.6	43.7	45	5	I25875	I25875 Sequence 4
23	16.6	43.7	45	5	I26268	I26268 Sequence 4
24	16.4	43.2	60	5	A17027	A17027 oligonucleo
25	16.4	43.2	60	5	A17427	A17427 oligonucleo
26	16.4	43.2	60	5	A21213	A21213 Synthetic n
27	16.4	43.2	60	5	A23859	A23859 oligonucleo
28	16.4	43.2	60	5	AR014402	AR014402 Sequence
29	16.4	43.2	60	5	AR059804	AR059804 Sequence
30	16.4	43.2	60	5	AR060211	AR060211 Sequence
31	16.4	43.2	60	5	I11845	I11845 Sequence 11
32	16.4	43.2	81	3	PIGPROG	M25036 Hog progast
33	16.4	43.2	92	10	HUMPHOS15	M59733 Human muscl
34	16.2	42.6	39	5	AR012248	AR012248 Sequence
35	16	42.1	39	5	I14946	I14946 Sequence 32
36	16	42.1	39	5	I73666	I73666 Sequence 32
37	16	42.1	73	4	CHKCSRC3	M61222 G.gallus c-
38	15.8	41.6	45	5	A47907	A47907 Sequence 4
39	15.8	41.6	45	5	AR028430	AR028430 Sequence
40	15.8	41.6	60	2	CTB12AOMP1	AF070239 Chlamydia
41	15.8	41.6	60	2	CTB12AOMP1	AF070239 Chlamydia
42	15.8	41.6	60	2	CTB12AOMP1	AF070239 Chlamydia
43	15.8	41.6	60	2	CTB12AOMP1	AF070239 Chlamydia
44	15.8	41.6	60	2	CTB12AOMP1	AF070239 Chlamydia
45	15.8	41.6	60	2	CTB12AOMP1	AF070239 Chlamydia

#### ALIGNMENTS



Best Local Similarity 68.6%; Pred. No. 3.6e+03;  
Matches 24; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2 atgtcgcagcttgaccatttggcaccacacatg 36  
|||||  
Db 26 ATGCTGCAGCTCCATCCACTCTTACAGCAGAGAG 60

## RESULT 4

CTB24AOMP1

LOCUS

DEFINITION Chlamydia trachomatis strain B24a major outer membrane protein

(omp-1) gene, partial cds.

ACCESSION AF070275.1 GI:3283544

VERSION

KEYWORDS

SEGMENT

SOURCE

ORGANISM

REFERENCE 1 (bases 1 to 60)

AUTHORS Hsieh, Y.-H. and Bobo, L.D.

TITLE Diversity of major outer membrane protein (omp-1) of Chlamydia trachomatis in trachoma endemic villages, Kongwa, Tanzania

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 60)

AUTHORS Hsieh, Y.-H. and Bobo, L.D.

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1998) Division of Pediatric Infectious Diseases, School of Medicine, Johns Hopkins University, Blalock 1138, 600 N. Wolfe Street, Baltimore, MD 21287-4933, USA

## FEATURES

Source

1. 60

/organism="Chlamydia trachomatis"

/strain="B24a"

/db\_xref="taxon:813"

&lt;1..&gt;60

/gene="omp-1"

/codon\_start=1

/transl\_table=1

/product="major outer membrane protein"

/protein\_id="AAC25245.1"

/db\_xref="GI:3283549"

/translation="AKPTATTGNAAPSTLTARE"

BASE COUNT

18 a 20 c 12 g 10 t

Query Match 45.8%; Score 17.4; DB 2; Length 60;

Best Local Similarity 68.6%; Pred. No. 3.6e+03;

Matches 24; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2 atgtcgcagcttgaccatttggcaccacacatg 36  
|||||  
Db 26 ATGCTGCAGCTCCATCCACTCTTACAGCAGAGAG 60

## RESULT 5

CTB24AOMP1

LOCUS

DEFINITION Chlamydia trachomatis strain B24a major outer membrane protein

(omp-1) gene, partial cds.

ACCESSION AF070255.1 GI:3283499

VERSION

KEYWORDS

SEGMENT

SOURCE

ORGANISM

REFERENCE 1 (bases 1 to 60)

AUTHORS Hsieh, Y.-H. and Bobo, L.D.

TITLE Diversity of major outer membrane protein (omp-1) of Chlamydia trachomatis in trachoma endemic villages, Kongwa, Tanzania

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 60)

AUTHORS Hsieh, Y.-H. and Bobo, L.D.

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1998) Division of Pediatric Infectious Diseases, School of Medicine, Johns Hopkins University, Blalock 1138, 600 N. Wolfe Street, Baltimore, MD 21287-4933, USA

## FEATURES

Source

1. 60

/organism="Chlamydia trachomatis"

/strain="B24a"

/db\_xref="taxon:813"

&lt;1..&gt;60

/gene="omp-1"

/codon\_start=1

/transl\_table=1

/product="major outer membrane protein"

/protein\_id="AAC25225.1"

/db\_xref="GI:3283504"

/translation="AKPTATTGNAAPSTLTARE"

BASE COUNT

18 a 20 c 12 g 10 t

Query Match 45.8%; Score 17.4; DB 2; Length 60;  
Best Local Similarity 68.6%; Pred. No. 3.6e+03;  
Matches 24; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2 atgtcgcagcttgaccatttggcaccacacatg 36  
|||||  
Db 26 ATGCTGCAGCTCCATCCACTCTTACAGCAGAGAG 60

## RESULT 6

CTB24BOMP1

LOCUS

DEFINITION Chlamydia trachomatis strain B24b major outer membrane protein

(omp-1) gene, partial cds.

ACCESSION AF070271.1 GI:3283535

VERSION

KEYWORDS

SEGMENT

SOURCE

ORGANISM

REFERENCE 1 (bases 1 to 60)

AUTHORS Hsieh, Y.-H. and Bobo, L.D.

TITLE Diversity of major outer membrane protein (omp-1) of Chlamydia trachomatis in trachoma endemic villages, Kongwa, Tanzania

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 60)

AUTHORS Hsieh, Y.-H. and Bobo, L.D.

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1998) Division of Pediatric Infectious Diseases, School of Medicine, Johns Hopkins University, Blalock 1138, 600 N. Wolfe Street, Baltimore, MD 21287-4933, USA

## FEATURES

Source

1. 60

/organism="Chlamydia trachomatis"

/strain="B24b"

/db\_xref="taxon:813"

&lt;1..&gt;60

/gene="omp-1"

/codon\_start=1

/transl\_table=1

/product="major outer membrane protein"

/protein\_id="AAC25241.1"

/db\_xref="GI:3283540"

/translation="AKPTATTGNAAPSTLTARE"

BASE COUNT

18 a 20 c 12 g 10 t

Query Match 45.8%; Score 17.4; DB 2; Length 60;  
Best Local Similarity 68.6%; Pred. No. 3.6e+03;  
Matches 24; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 atgtcgagcttgaccatttggcaccacatag 36  
|||||  
Db 26 ATGCTGCAGCTCCACTCTTACAGCAAGAG 60

RESULT 7  
CTB2ACMP1 60 bp DNA BCT 02-JUL-1998  
DEFINITION Chlamydia trachomatis strain B2a major outer membrane protein  
(omp-1) gene, partial cds.  
ACCESSION AF070251  
VERSION AF070251.1 GI:3283490  
KEYWORDS  
SEGMENT  
SOURCE  
ORGANISM  
1 of 4  
Chlamydia trachomatis.  
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
REFERENCE  
AUTHORS Hsieh, Y.-H. and Bobo, L.D.  
TITLE Diversity of major outer membrane protein (omp-1) of Chlamydia trachomatis in trachoma endemic villages, Kongwa, Tanzania  
JOURNAL Unpublished  
REFERENCE  
AUTHORS Hsieh, Y.-H. and Bobo, L.D.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-1998) Division of Pediatric Infectious Diseases, School of Medicine, Johns Hopkins University, Blalock 1138, 600 N. Wolfe Street, Baltimore, MD 21287-4933, USA

FEATURES  
source  
1. 60  
/organism="Chlamydia trachomatis"  
/strain="B2a"  
/db\_xref="taxon:813"  
<1. .>60  
/gene="omp-1"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein"  
/protein\_id="AAC25221.1"  
/db\_xref="GI:3283495"  
/translation="AKPYATGNAAPSTLTARE"  
BASE COUNT 18 a 20 c 12 g 10 t  
ORIGIN

Query Match 45.8%; Score 17.4; DB 2; Length 60;  
Best Local Similarity 68.6%; Pred. No. 3.6e+03;  
Matches 24; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 atgtcgagcttgaccatttggcaccacatag 36  
|||||  
Db 26 ATGCTGCAGCTCCACTCTTACAGCAAGAG 60

RESULT 8  
CTB2BOMP1 60 bp DNA BCT 02-JUL-1998  
DEFINITION Chlamydia trachomatis strain B2b major outer membrane protein  
(omp-1) gene, partial cds.  
ACCESSION AF070259  
VERSION AF070259.1 GI:3283508  
KEYWORDS  
SEGMENT  
SOURCE  
ORGANISM  
1 of 4  
Chlamydia trachomatis.  
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
REFERENCE  
AUTHORS Hsieh, Y.-H. and Bobo, L.D.  
TITLE Diversity of major outer membrane protein (omp-1) of Chlamydia

trachomatis in trachoma endemic villages, Kongwa, Tanzania  
Unpublished  
2 (bases 1 to 60)  
Hsieh, Y.-H. and Bobo, L.D.  
Direct Submission  
JOURNAL Submitted (02-JUN-1998) Division of Pediatric Infectious Diseases, School of Medicine, Johns Hopkins University, Blalock 1138, 600 N. Wolfe Street, Baltimore, MD 21287-4933, USA

FEATURES  
source  
1. 60  
/organism="Chlamydia trachomatis"  
/strain="B2b"  
/db\_xref="taxon:813"  
<1. .>60  
/gene="omp-1"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein"  
/protein\_id="AAC25223.1"  
/db\_xref="GI:3283513"  
/translation="AKPYATGNAAPSTLTARE"  
BASE COUNT 18 a 20 c 12 g 10 t  
ORIGIN

Query Match 45.8%; Score 17.4; DB 2; Length 60;  
Best Local Similarity 68.6%; Pred. No. 3.6e+03;  
Matches 24; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 atgtcgagcttgaccatttggcaccacatag 36  
|||||  
Db 26 ATGCTGCAGCTCCACTCTTACAGCAAGAG 60

RESULT 9  
CTB2COMP1 60 bp DNA BCT 02-JUL-1998  
DEFINITION Chlamydia trachomatis strain B2c major outer membrane protein  
(omp-1) gene, partial cds.  
ACCESSION AF070263  
VERSION AF070263.1 GI:3283517  
KEYWORDS  
SEGMENT  
SOURCE  
ORGANISM  
1 of 4  
Chlamydia trachomatis.  
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
REFERENCE  
AUTHORS Hsieh, Y.-H. and Bobo, L.D.  
TITLE Diversity of major outer membrane protein (omp-1) of Chlamydia trachomatis in trachoma endemic villages, Kongwa, Tanzania  
JOURNAL Unpublished  
REFERENCE  
AUTHORS Hsieh, Y.-H. and Bobo, L.D.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-1998) Division of Pediatric Infectious Diseases, School of Medicine, Johns Hopkins University, Blalock 1138, 600 N. Wolfe Street, Baltimore, MD 21287-4933, USA

FEATURES  
source  
1. 60  
/organism="Chlamydia trachomatis"  
/strain="B2c"  
/db\_xref="taxon:813"  
<1. .>60  
/gene="omp-1"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein"  
/protein\_id="AAC25233.1"  
/db\_xref="GI:3283523"  
/translation="AKPYATGNAAPSTLTARE"  
BASE COUNT 18 a 20 c 12 g 10 t  
ORIGIN



Query Match 45.8%; Score 17.4; DB 2; Length 60;  
Best Local Similarity 68.6%; Pred. No. 3.6e+03;  
Matches 24; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 2 atgtctgacgttgaccactttggcaccacacatag 36  
|||||  
Db 26 ATGCTGCAGCTCCATCCACTTACAGCAGAGAG 60

RESULT 10  
CTB2DOMP1 60 bp DNA BCT 02-JUL-1998  
DEFINITION Chlamydia trachomatis strain B2d major outer membrane protein  
(omp-1) gene, partial cds.  
ACCESSION AF070267  
VERSION AF070267.1 GI:3283526  
KEYWORDS  
SEGMENT  
SOURCE 1 of 4  
ORGANISM Chlamydia trachomatis.  
REFERENCE Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
AUTHORS 1 (bases 1 to 60)  
TITLE Hsieh,Y.-H. and Bobo,L.D.  
JOURNAL Diversity of major outer membrane protein (omp-1) of Chlamydia  
REFERENCE Unpublished  
AUTHORS 2 (bases 1 to 60)  
TITLE Hsieh,Y.-H. and Bobo,L.D.  
JOURNAL Submitted (02-JUN-1998) Division of Pediatric Infectious Diseases,  
School of Medicine, Johns Hopkins University, Balloock 1138, 600 N.  
Wolfe Street, Baltimore, MD 21287-4933, USA  
FEATURES  
source  
1. 60  
/organism="Chlamydia trachomatis"  
/strain="B2d"  
/db\_xref="taxon:813"  
<1. >60  
/gene="omp-1"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein"  
/protein\_id="AAC25237.1"  
/db\_xref="GI:3283531"  
/translation="AKPTATGTGNAAPSTLTARE"  
BASE COUNT 18 a 20 c 12 g 10 t  
ORIGIN

Query Match 45.8%; Score 17.4; DB 2; Length 60;  
Best Local Similarity 68.6%; Pred. No. 3.6e+03;  
Matches 24; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 2 atgtctgacgttgaccactttggcaccacacatag 36  
|||||  
Db 26 ATGCTGCAGCTCCATCCACTTACAGCAGAGAG 60

RESULT 11  
CTB2EOMP1 60 bp DNA BCT 02-JUL-1998  
DEFINITION Chlamydia trachomatis strain B2e major outer membrane protein  
(omp-1) gene, partial cds.  
ACCESSION AF070279  
VERSION AF070279.1 GI:3283553  
KEYWORDS  
SEGMENT  
SOURCE 1 of 4  
ORGANISM Chlamydia trachomatis.  
REFERENCE Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
AUTHORS 1 (bases 1 to 60)  
TITLE Hsieh,Y.-H. and Bobo,L.D.

TITLE Diversity of major outer membrane protein (omp-1) of Chlamydia  
trachomatis in trachoma endemic villages, Kongwa, Tanzania  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 60)  
AUTHORS Hsieh,Y.-H. and Bobo,L.D.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-1998) Division of Pediatric Infectious Diseases,  
School of Medicine, Johns Hopkins University, Balloock 1138, 600 N.  
Wolfe Street, Baltimore, MD 21287-4933, USA  
FEATURES  
source  
1. 60  
/organism="Chlamydia trachomatis"  
/strain="B2e"  
/db\_xref="taxon:813"  
<1. >60  
/gene="omp-1"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein"  
/protein\_id="AAC25249.1"  
/db\_xref="GI:3283558"  
/translation="AKPTATGTGNAAPSTLTARE"  
BASE COUNT 18 a 20 c 12 g 10 t  
ORIGIN

Query Match 45.8%; Score 17.4; DB 2; Length 60;  
Best Local Similarity 68.6%; Pred. No. 3.6e+03;  
Matches 24; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 2 atgtctgacgttgaccactttggcaccacacatag 36  
|||||  
Db 26 ATGCTGCAGCTCCATCCACTTACAGCAGAGAG 60

RESULT 12  
CTB2FOMP1 60 bp DNA BCT 02-JUL-1998  
DEFINITION Chlamydia trachomatis strain B2f major outer membrane protein  
(omp-1) gene, partial cds.  
ACCESSION AF070283  
VERSION AF070283.1 GI:3283562  
KEYWORDS  
SEGMENT  
SOURCE 1 of 4  
ORGANISM Chlamydia trachomatis.  
REFERENCE Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
AUTHORS 1 (bases 1 to 60)  
TITLE Hsieh,Y.-H. and Bobo,L.D.  
JOURNAL Diversity of major outer membrane protein (omp-1) of Chlamydia  
trachomatis in trachoma endemic villages, Kongwa, Tanzania  
REFERENCE Unpublished  
AUTHORS 2 (bases 1 to 60)  
TITLE Hsieh,Y.-H. and Bobo,L.D.  
JOURNAL Submitted (02-JUN-1998) Division of Pediatric Infectious Diseases,  
School of Medicine, Johns Hopkins University, Balloock 1138, 600 N.  
Wolfe Street, Baltimore, MD 21287-4933, USA  
FEATURES  
source  
1. 60  
/organism="Chlamydia trachomatis"  
/strain="B2f"  
/db\_xref="taxon:813"  
<1. >60  
/gene="omp-1"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein"  
/protein\_id="AAC25253.1"  
/db\_xref="GI:3283567"  
/translation="AKPTATGTGNAAPSTLTARE"  
BASE COUNT 18 a 20 c 12 g 10 t  
ORIGIN

Query Match 45.8%; Score 17.4; DB 2; Length 60;  
Best Local Similarity 68.6%; Pred. No. 3.6e+03;  
Matches 24; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 atgctgcagcttgaccatttggcaccacaatag 36  
|||||  
Db 26 ATGCTGCAGCTTCATCCACTTACGACGAAGAG 60

RESULT 13  
CTB3AOMP1 60 bp DNA BCT 02-JUL-1998  
LOCUS Chlamydia trachomatis strain B3a major outer membrane protein  
DEFINITION (comp-1) gene, partial cds.  
ACCESSION AF070287  
VERSION AF070287.1 GI:3283571  
KEYWORDS 1 of 4  
SEGMENT Chlamydia trachomatis.  
SOURCE Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
ORGANISM  
REFERENCE 1 (bases 1 to 60)  
AUTHORS Hsieh, Y.-H. and Bobo, L.D.  
TITLE Diversity of major outer membrane protein (omp-1) of Chlamydia trachomatis in trachoma endemic villages, Kongwa, Tanzania  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 60)  
AUTHORS Hsieh, Y.-H. and Bobo, L.D.  
TITLE Diversity of major outer membrane protein (omp-1) of Chlamydia trachomatis in trachoma endemic villages, Kongwa, Tanzania  
JOURNAL Unpublished  
REFERENCE Submitted (02-JUN-1998) Division of Pediatric Infectious Diseases, School of Medicine, Johns Hopkins University, Blalock 1138, 600 N. Wolfe Street, Baltimore, MD 21287-4933, USA  
FEATURES  
source 1. 60  
/organism="Chlamydia trachomatis"  
/strain="B3a"  
/db\_xref="taxon:813"  
<1..>60  
/gene="omp-1"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein"  
/protein\_id="AAC25257.1"  
/db\_xref="GI:3283576"  
/translation="AKPRATGNAAPSTLRARE"  
BASE COUNT 18 a 20 c 12 g 10 t  
ORIGIN

Query Match 45.8%; Score 17.4; DB 2; Length 60;  
Best Local Similarity 68.6%; Pred. No. 3.6e+03;  
Matches 24; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 atgctgcagcttgaccatttggcaccacaatag 36  
|||||  
Db 26 ATGCTGCAGCTTCATCCACTTACGACGAAGAG 60

RESULT 14  
CTB4AOMP1 60 bp DNA BCT 02-JUL-1998  
LOCUS Chlamydia trachomatis strain B4a major outer membrane protein  
DEFINITION (comp-1) gene, partial cds.  
ACCESSION AF070291  
VERSION AF070291.1 GI:3283580  
KEYWORDS 1 of 4  
SEGMENT Chlamydia trachomatis.  
SOURCE Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
ORGANISM  
REFERENCE 1 (bases 1 to 60)

AUTHORS Hsieh, Y.-H. and Bobo, L.D.  
TITLE Diversity of major outer membrane protein (omp-1) of Chlamydia trachomatis in trachoma endemic villages, Kongwa, Tanzania  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 60)  
AUTHORS Hsieh, Y.-H. and Bobo, L.D.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-1998) Division of Pediatric Infectious Diseases, School of Medicine, Johns Hopkins University, Blalock 1138, 600 N. Wolfe Street, Baltimore, MD 21287-4933, USA  
FEATURES  
source 1. 60  
/organism="Chlamydia trachomatis"  
/strain="B4a"  
/db\_xref="taxon:813"  
<1..>60  
/gene="omp-1"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein"  
/protein\_id="AAC25261.1"  
/db\_xref="GI:3283585"  
/translation="AKPRATGNAAPSTLRARE"  
BASE COUNT 18 a 20 c 12 g 10 t  
ORIGIN

Query Match 45.8%; Score 17.4; DB 2; Length 60;  
Best Local Similarity 68.6%; Pred. No. 3.6e+03;  
Matches 24; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 atgctgcagcttgaccatttggcaccacaatag 36  
|||||  
Db 26 ATGCTGCAGCTTCATCCACTTACGACGAAGAG 60

RESULT 15  
CTB4BOMP1 60 bp DNA BCT 02-JUL-1998  
LOCUS Chlamydia trachomatis strain B4b major outer membrane protein  
DEFINITION (comp-1) gene, partial cds.  
ACCESSION AF070295  
VERSION AF070295.1 GI:3283589  
KEYWORDS 1 of 4  
SEGMENT Chlamydia trachomatis.  
SOURCE Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
ORGANISM  
REFERENCE 1 (bases 1 to 60)  
AUTHORS Hsieh, Y.-H. and Bobo, L.D.  
TITLE Diversity of major outer membrane protein (omp-1) of Chlamydia trachomatis in trachoma endemic villages, Kongwa, Tanzania  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 60)  
AUTHORS Hsieh, Y.-H. and Bobo, L.D.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-1998) Division of Pediatric Infectious Diseases, School of Medicine, Johns Hopkins University, Blalock 1138, 600 N. Wolfe Street, Baltimore, MD 21287-4933, USA  
FEATURES  
source 1. 60  
/organism="Chlamydia trachomatis"  
/strain="B4b"  
/db\_xref="taxon:813"  
<1..>60  
/gene="omp-1"  
/codon\_start=1  
/transl\_table=11  
/product="major outer membrane protein"  
/protein\_id="AAC25265.1"  
/db\_xref="GI:3283594"  
/translation="AKPRATGNAAPSTLRARE"  
BASE COUNT 18 a 20 c 12 g 10 t  
ORIGIN

ORIGIN

Query Match 45.8%; Score 17.4; DB 2; Length 60;  
 Best Local Similarity 68.6%; Pred. No. 3.6e+03;  
 Matches 24; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2 atgtgcagcttgaccaattttggcaccacacatag 36  
 ||||||||||| | | | | | | | | | | |  
 Db 26 ATGCTGCAGCTCCATCCACTCTTACAGCAAGAGAG 60

Search completed: June 4, 2000, 16:05:44  
 Job time: 27896 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:24:20 ; Search time 322.35 Seconds  
(without alignments)  
29.494 Million cell updates/sec

Title: US-09-164-714-18

Sequence: 1 catgctgcagcttgcacatttggcacacaacataaggg 38

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 431286

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	19.6	51.6	96	1	Q40167	Sequence encoding
C 2	19.6	51.6	96	1	Q46018	CDNA derived from
C 3	19.6	51.6	96	1	Q47688	Encodes antiviral
C 4	19.6	51.6	96	1	Q56287	Epstein-Barr virus
C 5	19.6	51.6	96	1	Q75265	5'-terminal of ins
6	17.4	45.8	60	1	N92143	Chlamydia trachoma
7	17.4	45.8	60	1	N92147	Chlamydia trachoma
8	17.4	45.8	60	1	N97061	Sequence of C. tra
9	17.4	45.8	60	1	N97065	Sequence of C. tra
10	17.4	45.8	63	1	T28244	Feline herpesvirus
11	16.8	44.2	33	1	V45319	Heavy chain forwar
12	16.6	43.7	45	1	Q85044	3-Hydroxysteroid-o
13	16.6	43.7	45	1	T62671	Streptomyces sp. s
14	16.6	43.7	45	1	V11816	Human granulocyte
15	16.4	43.2	41	1	X19304	Feline herpesvirus
16	16.4	43.2	90	1	T28251	Human granulocyte
17	16.2	42.6	41	1	X19305	Human granulocyte
18	15.8	41.6	24	1	X22061	PCR primer for C.
C 19	15.8	41.6	31	1	X38735	Human genomic DNA
C 20	15.8	41.6	45	1	T10346	CAEV env gene TM f
C 21	15.8	41.6	60	1	N92131	Chlamydia trachoma
22	15.8	41.6	60	1	N92139	Sequence of Chlamy
23	15.8	41.6	60	1	N97049	Sequence of C. tra
24	15.8	41.6	60	1	N97057	Sequence of C. tra
25	15.8	41.6	60	1	X18016	C. trachomatis ser
26	15.8	41.6	60	1	X18008	C. trachomatis ser
27	15.8	41.6	61	1	N92127	Sequence of Chlamy
28	15.8	41.6	61	1	N97045	Sequence of C. tra
29	15.6	41.1	50	1	Q34114	Downstream sequenc
C 30	15.6	41.1	50	1	V76355	Staphylococcus aur
C 31	15.6	41.1	70	1	X16757	Alpha-helical peptid
32	15.6	41.1	71	1	V76550	Staphylococcus aur
33	15.4	40.5	50	1	Q12961	Template linker pr
34	15.4	40.5	50	1	T74411	HIV template linke

35	15.4	40.5	63	1	T28230	Feline herpesvirus
36	15.2	40.0	29	1	V43049	5' PCR primer Deg-
C 37	15.2	40.0	49	1	V13138	Oligonucleotide pr
38	15.2	40.0	51	1	T05046	LK26 heavy chain v
C 39	15.2	40.0	85	1	Q27291	TGF-alpha construc
C 40	15.2	40.0	85	1	Q44603	Plasmid pBS114 con
41	15.2	40.0	86	1	Q27293	TGF-alpha construc
42	15.2	40.0	86	1	Q44605	Plasmid pBS114 con
C 43	15	39.5	34	1	Q71835	Primer FIB2. fibro
C 44	15	39.5	34	1	T18304	Primer for amplifi
C 45	15	39.5	85	1	X04417	Fragment of human

## ALIGNMENTS

RESULT 1	
ID Q40167/c	Standard: DNA; 96 BP.
AC Q40167;	
DT 09-AUG-1993 (first entry)	
DE Sequence encoding 5' end of peptide with antiviral activity.	
KW HIV-1; HIV-2; HHV-6; therapy; antiviral; AIDS; ss.	
OS Sus scrofa domestica.	
PN EP-537722-A.	
PD 21-APR-1993.	
PF 14-OCT-1992; 117569.	
PR 15-OCT-1991; US-780084.	
PR 04-DEC-1991; US-803844.	
PA (KREM-) KREMERS-URBAN CO.	
PI Gawish AAS, Hermann WJ, Lawyer CH, Steinbach T,	
PI Tanaka SK, Magle SS;	
DR WPI; 93-127605/16.	
PT Polypeptide obtained from a porcine liver extract - useful for	
PT treating HIV-1, HIV-2, HIV-6 and AIDS infections	
PS Claim 6; Page 25; 42pp; English.	
CC The inventors claim a novel polypeptide with antiviral activity and	
CC a mol. wt. of 5-49 KD obtd. from an acetone-insoluble liver extract.	
CC The polypeptide has 5' and 3' amino acid sequences selected from	
CC those given in the specification encoded by DNA sequences also given	
CC in the specification. Synthetic peptides are also claimed. The	
CC polypeptides are used in compns. for the treatment of viral	
CC infections, partic. HIV-1, HIV-2 and human herpes virus (HHV-6)	
CC infections. The polypeptides are capable of inhibiting the	
CC replication and pathogenesis of the viruses.	
CC Sequence 96 BP; 22 A; 21 C; 23 G; 30 T;	
QY	
2 atgctgcagcttgcacatttggcacacaata 35	
Db 79 AAGCTGCAGCTTGACGTATTATTCACCTTGATA 46	
Query Match	51.6%; Score 19.6; DB 1; Length 96;
Best Local Similarity	73.5%; Pred. No. 21;
Matches 25; Conservative	0; Mismatches 9; Indels 0; Gaps 0;
RESULT 2	
ID Q46018/c	Standard: cDNA; 96 BP.
AC Q46018;	
DT 24-JAN-1994 (first entry)	
DE cDNA derived from liver, encoding therapeutic peptide.	
KW Alzheimers disease; dementia; medicament; chronic fatigue syndrome;	
KW peptide; Kutapressin; ss.	
OS Sus scrofa.	
PN EP-554887-A.	
PD 11-AUG-1993.	
PF 05-FEB-1993; 101819.	
PR 05-FEB-1992; US-835029.	
PA (KREM-) KREMERS-URBAN CO.	
PI Gawish AAS, Hermann WJ, Lawyer CH, Steinbach T;	
PI Magle SS;	



PT extract which is heat stable and insoluble in acetone but  
 PS soluble in water  
 Example 4; Columns 13-14; 25pp: English.  
 CC Polypeptides purified from mammalian liver extract which showed  
 CC cell protection activity when tested for anti-viral activity were  
 CC partially sequenced. A degenerate oligonucleotide primer (075269)  
 CC was designed based on the similarities between the different peptide  
 CC sequences (R66498-R66501) from active polypeptides KU 10,214 and KU  
 CC 10,215. The degenerate primer was used for PCR on cDNA prepared from  
 CC pig liver poly (A) + RNA. Amplified DNA fragments were analysed. One  
 CC insert of 500bp was found in six out of four randomly chosen isolates.  
 CC DNA sequencing determined 96bp at the 5'-end and 110bp at the 3'-end  
 CC of the insert (see 075265 and 075266, respectively). The 3'-end had  
 CC termination codons in all three reading frames giving 3 alternative  
 CC C-terminal sequences (i.e. Leu, R66494 or R66495). The active  
 CC polypeptides are thus characterised by sequences R66496 or R66497.  
 CC Sequence 96 BP; 22 A; 20 C; 24 G; 30 T

Query Match	51.6%	Score 19.6	DB 1	Length 96
Best Similarity	73.5%	Pred. No. 21		
Local Similarity	73.5%	Pred. No. 21		
Matches 25	Conservative 0	Mismatches 9	Indels 0	Gaps 0

QY 2 atgctcagccttgaccacatttggaccacacata 35  
| | | | | | | | | | | | | | | | | | | |  
79 AAGCTGCAGCTTGACGATATTATTCACACTTGATA 46  
| | | | | | | | | | | | | | | | | | | |

PT Chlamydia trachomatis genes - used for determ. of nucleotide and amino  
 PT sequences of the variable domains of the major outer membrane proteins.  
 PS Disclosure: Fig. 9; 49pp; English.  
 CC The major neutralising and serotyping antigenic determinants have been  
 CC located in the variable domain. The sequence was determined using  
 CC oligonucleotide primer extension sequencing. The variable domain is  
 CC recognised by MOMP specific monoclonal antibodies. The domain will be  
 CC useful in the construction of synthetic peptides, subunits or recombinant  
 CC chlamydial vaccines. They will also allow prodn. of reagents and methods  
 CC for development of new diagnostic tests for C. trachomatis infections  
 CC and serological tests for serotyping. The domain is bases 256-315 of the  
 CC MOMP gene.  
 SQ Sequence 60 BP; 18 A; 19 C; 13 G; 10 T;

Query Match	Similarity	Score	DB 1	length
Best local	24	68.6%	Pred. No. 1.5e+02	
Matches	24	Conservative	0	Mismatches 11; Indels 0; Gaps 0
Qy	2	atgctgcagcttgaccacatttgcgaccacacatag	36	
Db	26	atgctgcagcttcattccacttgcacgacgaagag	60	

CC to be the location of antigenic determinants recognised by MOMP  
 CC specific monoclonal antibodies. The nucleotide, amino acid  
 CC sequences and hydrophobicity/charge value analyses will assist in  
 CC the selection of appropriate MOMP antigenic determinants to be used  
 CC in the construction of synthetic peptides, subunits or recombinant  
 CC chlamydial vaccines. This will allow the prodn. or reagents and  
 CC methodologies applicable in the development of new diagnostic tests  
 CC for serotyping.  
 SQ Sequence 60 BP; 18 A; 19 C; 13 G; 10 T;

Query Match 45.8%; Score 17.4; DB 1; Length 60;  
 Best Local Similarity 68.6%; Pred. No. 1.5e+02;  
 Matches 24; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 2 atgcctgacgttaccatttggcaccaacatag 36  
 |||||  
 DB 26 ATGCTGACACTCCATCTGATGACAGAGAGAG 60

## RESULT 9

ID N97065 standard; DNA: 60 BP.  
 AC N97065;  
 DT 06-MAR-1992 (first entry)  
 DE Sequence of C. trachomatis serovar L2 major outer membrane protein (MOMP)  
 DE variable domain (VD) gene L2-VDI base pairs 256-315  
 KW Chlamydia trachomatis; antigen; monoclonal antibody; vaccine;  
 KW diagnosis; serotyping; non-immunologic assay; ss.  
 OS Chlamydia trachomatis.  
 FH Key Location/Qualifiers  
 FT cds 1..60  
 FT US7324664-A.  
 PN 29-AUG-1989.  
 PF 17-MAR-1989; 324664.  
 PR 17-MAR-1989; US-324664.  
 PA (USSH ) US DEPT HEALTH & HUMAN.  
 PI Caldwell HD, Ying YX, Watkins NG;  
 DR WPI: 89-339697/46.  
 DR P-PSDB: p98430.

PT Chlamydia trachomatis genes - used for determ. of nucleotide and  
 PT amino sequences of the variable domains of the major outer  
 PT membrane proteins  
 PS Disclosure; Fig 10; 49pp; English.  
 CC The inventors sequenced the 4 MOMP VDS of ten C. trachomatis  
 CC serovars and the amino acid sequences were deduced. The MOMP VDS  
 CC with the greatest total hydrophobicity and charge values were found  
 CC to be the location of antigenic determinants recognised by MOMP  
 CC specific monoclonal antibodies. The nucleotide, amino acid  
 CC sequences and hydrophobicity/charge value analyses will assist in  
 CC the selection of appropriate MOMP antigenic determinants to be used  
 CC in the construction of synthetic peptides, subunits or recombinant  
 CC chlamydial vaccines. This will allow the prodn. or reagents and  
 CC methodologies applicable in the development of new diagnostic tests  
 CC for serotyping.  
 SQ Sequence 60 BP; 18 A; 19 C; 13 G; 10 T;

Query Match 45.8%; Score 17.4; DB 1; Length 60;  
 Best Local Similarity 68.6%; Pred. No. 1.5e+02;  
 Matches 24; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 2 atgcctgacgttaccatttggcaccaacatag 36  
 |||||  
 DB 26 ATGCTGACACTCCATCTGATGACAGAGAGAG 60

## RESULT 10

ID T28244 standard; DNA: 63 BP.  
 AC T28244;  
 DT 18-DEC-1996 (first entry)

DE Feline herpesvirus homology vector 669-42.04 junction H fragment.  
 KW Feline herpesvirus; FHV; deletion; insertion; glycoprotein E;  
 KW gp E; unique short region; attenuated virus; vaccine; protection;  
 KW cat; vector; delivery; antigen; homology vector; junction H; ds.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_feature 1..13  
 FT /tag= a  
 FT /label= pseudorabies\_virus (Bam HI #10)  
 FT misc\_feature 14..42  
 FT /tag= b  
 FT /label= synthetic\_linker  
 FT misc\_feature 43..63  
 FT /tag= c  
 FT /label= feline\_herpesvirus (Eco RI E)

PN W09613575-A1.  
 PD 09-MAY-1996.  
 PF 26-OCT-1995; U13975.  
 PR 26-OCT-1994; US-329883.  
 PA (SYTR ) SYNTRO CORP.  
 PI Cochran MD, McDonnell MW;  
 DR WPI: 96-239489/24.  
 PT Recombinant feline herpes virus attenuated by alteration of the ge  
 PT gene - and related homology vectors, useful in vaccines, also as  
 PT vectors for delivering antigens or therapeutic agents to mammals and  
 PT birds  
 PS Claim 49: Fig 12E; 162pp; English.  
 CC The present sequence is a junction H fragment from the claimed  
 CC feline herpesvirus (FHV) homology vector 669-42.04, which was  
 CC constructed for the purpose of deleting a portion of the FHV  
 CC glycoprotein E (gp E) gene coding region, and inserting a foreign  
 CC DNA. It incorporates an E. coli beta-galactosidase gene, and a  
 CC feline immunodeficiency virus (FIV) protease (gag) gene flanked by  
 CC FHV DNA. The foreign genes were inserted into the NotI site in the  
 CC FHV homology vector 644-09.04.  
 CC Altering the gp E gene of FHV, so that it no longer produces  
 CC functional gp E, attenuates the virus. The attenuated virus is  
 CC useful in vaccines to protect cats against FHV (and opt. against  
 CC other diseases if appropriate DNA is inserted), and as a vector  
 CC for delivering vaccinating antigens or therapeutic agents (e.g.  
 CC antisense mols., ribozymes, interferon inducers, hormones,  
 CC lymphokines, etc.) to other mammals (including humans) or birds.  
 SQ Sequence 63 BP; 11 A; 22 C; 20 G; 10 T;

Query Match 45.8%; Score 17.4; DB 1; Length 63;  
 Best Local Similarity 77.8%; Pred. No. 1.5e+02;  
 Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 4 gctgcagcttaccatttggcacca 30  
 |||||  
 DB 36 GCTGACAGTGCAGAGCTTCTGACACCA 62

## RESULT 11

ID V45319/C  
 AC V45319;  
 DT 27-OCT-1998 (first entry)

DE Heavy chain forward primer.  
 KW J chain; targeting molecule; epithelial; beta-sheet; asthma;  
 KW cancer; inflammatory disorder; autoimmune disorder; celiac disease;  
 KW colitis; pneumonia; cystic fibrosis; ss; PCR; primer; amplification.  
 OS Synthetic.  
 PN W09830592-A1.  
 PD 16-JUL-1998.  
 PF 09-JAN-1998; U00542.  
 PR 10-JAN-1997; US-782481.  
 PA (EPIC-) EPICYTE PHARM INC.  
 PI Fitchen JH, Hein MB, Hlatk AC;  
 DR WPI: 98-399067/34.  
 PT New epithelial tissue targeting agent - used to deliver  
 PT biologically active compounds to an epithelial surface for



```
PT Internalisation
CC Example 1: Page 61; 142pp; English.
CC The primers V45313-V45322 are used in the method of the invention
CC to synthesize targeting molecules (TM). The TMs are used to target
CC biological agents to epithelial surfaces at which they can be
CC internalised. The TMs comprise a polypeptide that: (a) forms a closed
CC covalent loop; (b) contains at least 3, preferably 4, peptide domains
CC having beta-sheet character separated by domains lacking beta-sheet
CC character; and (c) is not full length dimeric IgA. The TMs are useful to
CC prevent and/or treat diseases associated with epithelial surfaces, e.g.
CC asthma, cancer, (myco)bacterial, viral or fungal infection, inflammatory
CC disorders, autoimmune disorders, celiac disease, colitis, pneumonia
CC and cystic fibrosis.
SQ Sequence 33 BP; 7 A; 5 C; 9 G; 7 T;

Query Match 44.2%; Score 16.6; DB 1; Length 33;
Best Local Similarity 64.3%; Pred. No. 2.3e+02;
Matches 18; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 tgcgtcagcttgaccatttgcaccac 30
   |||||||:|||||:|||||
DB 28 TSCTGAGYTKSACCTGATCTGTACCA 1

RESULT 12
Q85044 085044 standard; DNA; 45 BP.
AC 085044;
DT 02-SEP-1995 (first entry)
DE 3-Hydroxysteroid-oxidase peptide N2-derived probe.
KM 3-Hydroxysteroid-oxidase: N2; DNA probe; insecticide;
KW transgenic plant; insect control; ss.
OS Synthetic.
PN W09501098-A.
PD 12-JUN-1995.
PF 24-JUN-1994; U07252.
PR 28-JUN-1993; US-083948.
PA (MONS ) MONSANTO CO.
PI Corbin DR, Greenplate JT, Jennings MG, Purcell JP.
PI Sammons RD;
DR WPI: 95-060750/08.
PT Controlling lepidopteran insect infestation of plants - by
PT providing a 3-hydroxy;steroid oxidase for ingestion by the insect.
PS Disclosure; Page 13; 39pp; English.
CC Hybridization probe N2 is used with probes C1 (085045) and C2
CC (085046) on Southern blots of Streptomyces sp. A19249 genomic DNA
CC for isolation and cloning of the 3-hydroxysteroid-oxidase (3-HSO)
CC gene. This sequence may be used to construct a cotton
CC (Gossypium hirsutum) or maize (Zea mays) transgenic plant
CC expressing 3-HSO. Alternatively, the sequence may be
CC expressed in plant-colonizing microorganisms or in Escherichia
CC coli, and applied directly to the plant, for controlling
CC lepidopteran insect infestations.
SQ Sequence 45 BP; 8 A; 14 C; 15 G; 8 T;

Query Match 43.7%; Score 16.6; DB 1; Length 45;
Best Local Similarity 71.0%; Pred. No. 3e+02;
Matches 22; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 atgctgagcttgaccatttgcaccac 32
   |||||||:|||||:|||||
DB 13 ATGCTGAGATGGCCAGCTGTGGAACACG 43

RESULT 13
T62671 13
ID T62671 standard; DNA; 45 BP.
AC T62671;
DT 22-MAY-1997 (first entry)
DE 3-hydroxysteroid oxidase gene probe N2.
KW Plant; insect; pest; control; insecticide; toxin; ingestion; probe;
```

```
KW 3-hydroxysteroid oxidase; boll weevil; lepidoptera; kill; stunt growth;
KM larvae; gut; ingestion; hybridisation; Streptomyces sp. A19249; ss.
OS Synthetic.
PN US5536862-A.
PD 24-SEP-1996.
PF 23-SEP-1991; 762682.
PR 23-SEP-1991; US-762682.
PR 04-SEP-1992; US-937195.
PR 28-JUN-1993; US-083948.
PR 24-FEB-1995; US-393785.
PR 07-JUN-1995; US-475694.
PA (MONS ) MONSANTO CO.
PI Corbin DR, Greenplate JT, Jennings MG, Purcell JP.
PI Sammons RD;
DR WPI: 96-442362/44.
PT Genetically transformed plant-colonising microorganism contg.
PT 3-hydroxy;steroid oxidase gene - used for the control of insects
PT esp. lepidopteran and boll weevil.
PS Disclosure; Column 13-14; 18pp; English.
CC Plant insect pests can be controlled by applying to the plant environment
CC or plant seed an insecticidally effective amt. of a protein toxin for
CC ingestion by the insect. In particular, a plant-colonising microorganism
CC genetically engineered to contain and express a heterologous gene
CC comprising a DNA sequence encoding 3-hydroxysteroid oxidase (3-HO) is
CC applied to the plant environment or seed. The protein toxin controls
CC insect infestation, especially boll weevil or lepidopteran. The protein
CC is believed to kill or stunt growth of boll weevil larvae and stunt the
CC growth of lepidopteran larvae by some action in the gut after ingestion.
CC T62671-73 are hybridisation probes used to identify the 3-HO gene in
CC Streptomyces sp. A19249. They were designed based on peptides from the
CC 3-HO protein.
SQ Sequence 45 BP; 8 A; 14 C; 15 G; 8 T;

Query Match 43.7%; Score 16.6; DB 1; Length 45;
Best Local Similarity 71.0%; Pred. No. 3e+02;
Matches 22; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 atgctgagcttgaccatttgcaccac 32
   |||||||:|||||:|||||
DB 13 ATGCTGAGATGGCCAGCTGTGGAACACG 43

RESULT 14
V11816 14
ID V11816 standard; DNA; 45 BP.
AC V11816;
DT 11-AUG-1998 (first entry)
DE Streptomyces sp. strain A19249 3-hydroxysteroid oxidase N2 probe.
KM 3-hydroxysteroid oxidase; lepidoptera; plant infestation; CryIA protein;
KW Bacillus thuringiensis; insecticide; cholesterol; crystal protein;
KW synergism; mortality; stunting; larvae; probe; ss.
OS Synthetic.
PN US763245-A.
PD 09-JUN-1998.
PF 10-SEP-1996; 712057.
PR 10-SEP-1996; US-712057.
PR 23-SEP-1991; US-762682.
PR 04-SEP-1992; US-937195.
PR 28-JUN-1993; US-083948.
PR 24-FEB-1995; US-393785.
PR 07-JUN-1995; US-475694.
PA (MONS ) MONSANTO CO.
PI Corbin DR, Greenplate JT, Pershing JC, Purcell JP.
PI WPI: 98-347327/30.
DR P-RSDB; W59120.
PT Insect control on plants - comprising administration of Bacillus
PT thuringiensis CryIA protein in combination with 3-hydroxy;steroid
PT oxidase
PS Disclosure; Column 29; 25pp; English.
CC This sequence is a hybridisation probe which is used in a method for
CC controlling lepidopteran infestation of plants. This method
```

CC involves the construction of a recombinant expression vector encoding a  
 CC 3-hydroxysteroid oxidase and a Bacillus thuringiensis CryIA protein  
 CC each sequence being operatively linked to a promoter to express the  
 CC nucleic acid sequences separately or in tandem. This construct can be  
 CC used to produce a genetically transformed plant that produces an  
 CC insecticidally effective amount of a CryIA protein and a 3-hydroxysteroid  
 CC oxidase. combinations of CryIA crystal proteins and cholesterol oxidase  
 CC have synergistically enhanced effects on mortality and stunting of  
 CC lepidopteran larvae.  
 SQ Sequence 45 BP; 8 A; 14 C; 15 G; 8 T;

Query Match 43.7%; Score 16.6; DB 1; Length 45;  
 Best Local Similarity 71.0%; Pred. No. 3e+02;  
 Matches 22; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 2 atgtcagcattgacacatttgccaccac 32  
 ||||| | | |||| | ||| |||||  
 DB 13 ATGCTGAGATGGCCGCTGTGAACACAGC 43

## RESULT 15

ID X19304 standard; DNA; 41 BP.  
 AC X19304;  
 DT 17-MAY-1999 (first entry)  
 DE Human granulocyte colony-stimulating factor sense primer oligo 8.  
 KW Human; granulocyte colony-stimulating factor; G-CSF; hg-CSF;  
 KW haematopoiesis disorder; primer; ss.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN WO9853072-A1.  
 PD 26-NOV-1998.  
 PF 21-MAY-1998; KR0125.  
 PR 22-MAY-1997; KR-020054.  
 PA (JELI-) JELI PHARM CO LTD.  
 PA (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOGY.  
 PI Choi B, Jun H, Lee Y, Sohn M;  
 PI WPI; 99-131691/11.  
 PT New human granulocyte colony-stimulating factor (hg-CSF) containing  
 PT a peptide at the N-terminus - useful in the treatment of  
 PT haematopoiesis disorders  
 PS Example 1; Fig 1; 19pp; English.  
 CC The present invention describes the N-terminal peptide of human  
 CC granulocyte colony-stimulating factor (hg-CSF). The present invention  
 CC also describes: (1) recombinant plasmid pYHM-G-CSF containing the cDNA  
 CC for the peptide; and (2) E. coli BL21 (pYHM-G-CSF) (KCTC 0477BP)  
 CC transformed with the plasmid. The plasmid and E. coli strain are used  
 CC to produce high yields of highly purified hg-CSF, which when refolded  
 CC has comparable biological activity of G-CSF. hg-CSF is used to treat  
 CC haematopoiesis disorders. The method of producing hg-CSF is less time-  
 CC consuming and, therefore, more economical than previous methods because  
 CC the protein is expressed in high yields without the need for further  
 CC processing steps. X19290 to X19315 represent primer used in the  
 CC synthesis of hg-CSF from an example of the present invention.  
 SQ Sequence 41 BP; 6 A; 14 C; 11 G; 10 T;

Query Match 43.2%; Score 16.4; DB 1; Length 41;  
 Best Local Similarity 67.6%; Pred. No. 3.5e+02;  
 Matches 23; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 4 gctgcagcttgacacatttgccacacatag 37  
 ||||| | | |||| | ||| |||||  
 DB 39 GCTGCAGTGTGTCCAAAGTGGACCACTCCGG 6

Search completed: June 4, 2000, 16:24:21  
 Job time: 28917 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 13:53:32 ; Search time 4521.53 Seconds  
(without alignments)  
34.064 Million cell updates/sec

Title: US-09-164-714-18

Perfect score: 38  
Sequence: 1 catgctgcagcttgaccacatttggcaccacacatagg 38

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026511650 residues

Total number of hits satisfying chosen parameters: 119968

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result

Query

SUMMARIES





Query Match 44.7%; Score 17; DB 35; Length 97;  
 Best Local Similarity 69.7%; Pred. No. 2.9e+03;  
 Matches 23; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 2 atgcctgacacttgacacatttggcaccacacat 34  
 1 |||||  
 Db 90 AGGCTGCAGTCGAGCCAGATGTCACACCCACAT 58

RESULT 5  
 R85175 92 bp mRNA EST 14-AUG-1995  
 LOCUS y043d03.r1 Soares adult brain N2b4HB55Y Homo sapiens cDNA clone  
 DEFINITION IMAGE:180677 5' similar to gb:W61764 TUBULIN GAMMA CHAIN (HUMAN);,  
 mRNA sequence.

ACCESSION R85175  
 VERSION R85175.1 GI:943581  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homidae; Homo.  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maier, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P., and  
 Wilson, R.

TITLE The WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu

Insert Size: 2081  
 High quality sequence starts: 1  
 High quality sequence stops: 1  
 Source: IMAGE Consortium, LLNL  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Trace considered overall poor quality  
 Insert Length: 2081 Std Error: 0.00  
 Seq primer: M13RP1  
 High quality sequence stop: 1.

FEATURES  
 SOURCE Location/Qualifiers  
 1..92  
 /organism="Homo sapiens"  
 /db\_xref="GDB:3827325"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:180677"  
 /clone\_lib="Soares adult brain N2b4HB55Y"  
 /sex="Male"  
 /dev\_stage="55-year old"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: brain; Vector: p773D (Pharmacia) with a  
 modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTTCACATCTGAGTGGAGCGCCGCTTTTCTTTTCTTTTCTT 3']  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified p773 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 53. Library constructed by Bento  
 Soares and M. Fatima Bonaldo. The adult brain RNA was  
 provided by Dr. Donald H. Gilden. Tissue was acquired  
 17-18 hours after death which occurred in consequence of a  
 ruptured aortic aneurysm. RNA was prepared from a pool of  
 tissues representing the following areas of the brain:  
 frontal, parietal, temporal and occipital cortex from the  
 left and right hemispheres, subcortical white matter,

BASE COUNT 18 a 24 c 26 g 18 t 6 others  
 ORIGIN  
 Query Match 44.2%; Score 16.8; DB 23; Length 92;  
 Best Local Similarity 64.9%; Pred. No. 3.5e+03;  
 Matches 24; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 1 catgctgacacttgacacatttggcaccacatag 37  
 1 |||||  
 Db 21 CACCCTGCAGCTTGCGCAGTCGACACGATTTGG 57

RESULT 6  
 A1938341/C 61 bp mRNA EST 02-AUG-1999  
 LOCUS sc43e05.y1 Gm-c1014 glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
 DEFINITION Gm-c1014-1881 5' similar to TR:Q42462 Q42462 ANT ;, mRNA sequence.

ACCESSION A1938341  
 VERSION A1938341  
 KEYWORDS EST.  
 SOURCE soybean.  
 ORGANISM Glycine max

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
 eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
 Glycine.  
 AUTHORS 1 (bases 1 to 61)  
 Shoemaker, R., Kelm, P., Vodkin, L., Erpelting, J., Corryell, V.,  
 Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J.,  
 Beck, C., Wylie, T., Underwood, K., Stephens, M., Theising, B., Allen, M.,  
 Bowers, Y., Person, B., Swaller, T., Gibson, M., Pape, D., Harvey, N.,  
 Schurr, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,  
 McCann, R., Waterston, R. and Wilson, R.

TITLE Public Soybean EST Project  
 JOURNAL Unpublished (1999)  
 COMMENT On Jun 5, 1998 this sequence version replaced g1:3188742.  
 Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World  
 Parkway Circle St. Louis, Missouri 63134 For further information  
 call: (800) 450-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
 427-3324 or contact: clones@genomesystems.com or  
 info@genomesystems.com web site: www.genomesystems.com  
 Trace considered overall poor quality  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: 40RP from Gibco  
 High quality sequence stop: 1.

FEATURES  
 SOURCE Location/Qualifiers  
 1..61  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1014-1881"  
 /clone\_lib="Gm-c1014"  
 /tissue\_type="leaves, 2-3 week old seedlings, greenhouse  
 grown"  
 /lab\_host="DH10B"  
 /note="Vector: p773Pac (p773, Pharmacia); Site:1: EcoRI;  
 Site:2: HindIII; This cDNA library was constructed from  
 mRNA isolated from leaves of 2-3 week old greenhouse grown  
 plants. Complementary DNA was synthesized from mRNA using  
 a 3' anchored poly (dT) primer. EcoRI adapters were  
 ligated to the blunt-ended cDNA fragments followed by  
 digestion with EcoRI and HindIII. The cDNA fragments were  
 directionally cloned into the EcoRI-HindIII restriction  
 site of the p773-Pac vector. The ligated cDNA fragments

were transformed into DH10B host cells (Gibco BRL). This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelting.

BASE COUNT 23 a 12 c 17 g 9 t

Query Match 42.1%; Score 16; DB 63; Length 61;  
Best Local Similarity 79.2%; Pred. No. 6.6e+03;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 tgcctgcagcttgaccattggc 26  
||||| ||||| ||||| |||||  
Db 49 TCCTGCACATTGGCCATTCTTGC 26

RESULT 7  
AA465907 67 bp mRNA EST 04-AUG-1997  
LOCUS V898404.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone  
DEFINITION IMAGE:833358 5', mRNA sequence.

ACCESSION AA465907  
VERSION AA465907.1 GI:2192047  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 67)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
COMMENT On Sep 29, 1993 this sequence version replaced gi:430089.  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:493574  
Seq primer: -40ml3 fwd. ET from Amersham.

## FEATURES

## Source

1..67  
/organism="Mus musculus"  
/strain="B6D2 F1/J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:833358"  
/clone\_lib="Knowles Solter mouse 2 cell"  
/tissue\_type="embryo"  
/dev\_stage="2-cell"  
/lab\_host="DH10B"  
/note="Organ: embryo; Vector: pBluescribe (modified);  
Site\_1: Muir, Site\_2: Salt; Cloned unidirectionally from  
mRNA prepared from 13,500 2-cell stage embryos. Primer:  
Salt(dT): 5'-CGGTGACGCGTCGACGCTTTTGT-3'. CDNA:  
were cloned into the Muir/Salt sites of a modified  
pBluescribe vector using commercial linkers (NEB).  
Average insert size: 1.2 kb."

BASE COUNT 11 a 24 c 13 g 19 t

## ORIGIN

Query Match 42.1%; Score 16; DB 34; Length 67;  
Best Local Similarity 68.8%; Pred. No. 6.8e+03;  
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 catctgcagcttgaccattggcaccaac 32  
||| ||||| ||| ||||| |||  
Db 13 CAGCTTCACGCTTGCCCTGTGTGGACCCAC 44

RESULT 8  
AA165762 79 bp mRNA EST 12-FEB-1997  
LOCUS ms60f11.r1 StrataGene mouse embryonic carcinoma (#937317) Mus  
DEFINITION musculus cDNA clone IMAGE:615981 5' similar to "R:E93245 E93245 ETN  
INSERT IN THE FAS APOPTOSIS GENE OF MRL-IPR/IPR. [1]"; mRNA  
sequence.

ACCESSION AA165762  
VERSION AA165762.1 GI:1743977  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 79)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1325676.  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:376805  
Seq primer: -28ml3 rev1 ET from Amersham  
High quality sequence stop: 75.

## FEATURES

## Source

1..79  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:615981"  
/clone\_lib="StrataGene mouse embryonic carcinoma  
(#937317)"  
/tissue\_type="carcinoma"  
/dev\_stage="embryonic"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; Cloned unidirectionally. Primer: Oligo dT. Pig cell  
line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'  
adaptor sequence: 5' GATTTCGACGACG 3' -3' adaptor  
sequence: 5' CTCGAGTGTGTGTGTGTGTGTGT 3'."

BASE COUNT 20 a 19 c 26 g 14 t

## ORIGIN

Query Match 42.1%; Score 16; DB 29; Length 79;  
Best Local Similarity 68.8%; Pred. No. 7.1e+03;  
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 7 gcagcttgaccattggcacacatagg 38  
||| ||| ||||| ||||| |||  
Db 25 GCGGCCGACACATTGTGGCCGACAACTGGG 56

RESULT 9  
W05206 97 bp mRNA EST 23-APR-1996  
LOCUS za42h07.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone  
DEFINITION IMAGE:295261 5' similar to PIR:S45038 S45038 Cabp1 protein - rat ;,





This clone is available royalty-free through LIND; contact the IMAGE Consortium (info@image.lind.gov) for further information.  
Insert Length: 2951 Std Error: 0.00  
Seq primer: -41ml3 fwd, ET from Amersham  
High quality sequence stop: 71.

## FEATURES

source

1. 84  
/organism="Homo sapiens"  
/db\_xref="GDB:5587628"  
/db\_xref="taxon:9606"  
/clone="IMAGE:684276"  
/clone\_lib="NCI\_CGAP\_GCB1"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(OT) primer  
[5'-TGTACCAATCTGAGTGGAGCGCCGCTCATTTTCTTTTCTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

27 a 8 c 14 g 35 t

ORIGIN

Query Match 41.6%; Score 15.8; DB 30; Length 84;  
Best Local Similarity 65.7%; Pred. No. 8.9e+03;  
Matches 23; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Db 69 CATTGCTTACTTGTGACATTTTGTGACTTCCAAA 35

RESULT 12

B32761 87 bp DNA GSS 17-OCT-1997

LOCUS

DEFINITION HS-1015-B2-F10-MR.ab1 CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate=CT 790 Col=20 Row=L, genomic survey sequence.

ACCESSION B32761

VERSION B32761.1 GI:2532130

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 87)

Mahairas G.G., Zackrone K.D., Smith T., Tipton S., Schmidt S.,

Tralcoff R., Adajian C., Blanchard A., West A. and Hood L.E.

Construction of a Characterized Clone Resource for Genomic

Sequencing: Generation and Preliminary Analysis of 20,000 Sequence

Tagged Connectors

Unpublished (1997)

Contact: Mahairas GG, Zackrone KD, Hood L

University of Washington

Seattle, WA 98195, USA

Tel: (206) 616-8744

Fax: (206) 685-7301

Email: kzackrone@u.washington.edu

Sequence Tagged Connector

Plate: CT 790 Row: L Column: 20

Class: BAC ends

High quality sequence stop: 87.

Location/Qualifiers

1. 87

/organism="Homo sapiens"

/db\_xref="taxon:9606"  
/clone="Plate=CT 790 Col=20 Row=L"  
/clone\_lib="CIT Human Genomic Sperm Library C"  
/sex="M"  
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT

20 a 19 c 12 g 36 t

ORIGIN

Query Match 41.6%; Score 15.8; DB 84; Length 87;  
Best Local Similarity 65.7%; Pred. No. 8.9e+03;  
Matches 23; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Db 53 CTTTCTGATCAGATCAATCTGCTCATATAAA 87

RESULT 13

A1155501 90 bp mRNA EST 11-FEB-1997

LOCUS

DEFINITION mr90a11.r1 StrataGene mouse embryonic carcinoma (#937317) Mus musculus cDNA clone IMAGE:604700 5' similar to TR:G1263081 G1263081 MARINER TRANSPOSASE. ; mRNA sequence.

ACCESSION A1155501

VERSION A1155501.1 GI:1727140

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 90)

Marras M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T.,

Geisel S., Kucaba T., Lacy M., Le M., Martin J., Morris M.,

Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B.,

Theising B., Wylie T., Lennon G., Soares B., Wilson R. and

Waterston R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

On Sep 12, 1996 this sequence version replaced gi:1315818.

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LIND; contact the

IMAGE Consortium (info@image.lind.gov) for further information.

Possible reversed clone: similarity on wrong strand

Seq primer: -28ml3 rev1 ET from Amersham

High quality sequence stop: 80.

FEATURES

source

1. 90  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:604700"  
/clone\_lib="Stratagene mouse embryonic carcinoma (#937317)"  
/tissue\_type="carcinoma"  
/dev\_stage="embryonic"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dT, P19 cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGACGAG 3' -3' adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3' "

BASE COUNT

30 a 17 c 20 g 23 t

ORIGIN

Query Match 41.6%; Score 15.8; DB 29; Length 90;  
 Best Local Similarity 65.7%; Pred. No. 9e+03;  
 Matches 23; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 3 tgcgcagcttgaccattggcaccacatagg 37  
 ||||| ||||| ||||| ||||| |||||  
 Db 81 TCGTGCTTATGACCCATTTAAACCCCAATTAAG 47

RESULT 14  
 A1475433 91 bp mRNA EST 09-MAR-1999  
 LOCUS C183d04.x1 NCI\_CGAP\_Col4 Homo sapiens cDNA clone IMAGE:2153671 3',  
 DEFINITION mRNA sequence.  
 ACCESSION A1475433  
 VERSION A1475433.1 GI:4328478  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 91)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 On Jun 22, 1998 this sequence version replaced gi:3247120.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskalkuk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)

Seq primer: -40UP from Gibco  
 High quality sequence stop: 85.

FEATURES  
 source  
 1..91  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2153671"  
 /clone\_lib="NCI\_CGAP\_Col4"  
 /tissue\_type="moderately-differentiated adenocarcinoma"  
 /lab\_host="DH10B"  
 /note="Organ: colon; Vector: pCMV-SPORT6; Site\_1: SalI;  
 Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.7 kb. Life Technologies catalog #: 11531-019"

BASE COUNT 8 a 5 c 20 g 58 t  
 ORIGIN

Query Match 41.6%; Score 15.8; DB 47; Length 91;  
 Best Local Similarity 65.7%; Pred. No. 9e+03; 12; Indels 0; Gaps 0;  
 Matches 23; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 4 gctgcagcttgaccattggcaccacatagg 38  
 ||||| ||||| ||||| ||||| |||||  
 Db 49 GCGGGGCTGTGTTTATTTTGGCCCAAAAAGG 83

RESULT 15  
 AA016655 100 bp mRNA EST 02-AUG-1996  
 LOCUS mg89h09.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA  
 DEFINITION clone IMAGE:440225 5' similar to SW:ATPM\_HUMAN Q06055 ATP SYNTHASE  
 LIPID-BINDING PROTEIN P2 PRECURSOR ;, mRNA sequence.  
 ACCESSION AA016655

VERSION AA016655.1 GI:1478885  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 100)  
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
 The WashU-HMIT Mouse EST Project  
 Unpublished (1996)  
 On Apr 14, 1993 this sequence version replaced gi:693329.  
 CONTACT: Marra M/Mouse EST Project  
 WashU-HMIT Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 MGI:265561  
 Trace considered overall poor quality  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -28M13 rev2 from Amersham  
 High quality sequence stop: 1.

FEATURES  
 source  
 1..100  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:440225"  
 /clone\_lib="Soares mouse embryo NBME13.5 14.5"  
 /sex="unknown"  
 /tissue\_type="embryo"  
 /dev\_stage="13.5-14.5dpc total fetus"  
 /lab\_host="DH10B"  
 /note="Vector: p773D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5',  
 TGTTCACATCTGAGATGGAGCGCGCGAATATTTTTTTTTTTTTTTT  
 T 3'], on equal amounts of mRNA from 2 13.5dpc and 2  
 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne  
 State Univ., from 2 ]; double-stranded cDNA was ligated to  
 Eco RI adaptors (Pharmacia), digested with Not I and  
 cloned into the Not I and Eco RI sites of the modified  
 p7733 vector. Library went through one round of  
 normalization, and was constructed by Bento Soares and  
 M.Fatima Bonaldo."

BASE COUNT 20 a 33 c 23 g 24 t  
 ORIGIN

Query Match 41.6%; Score 15.8; DB 27; Length 100;  
 Best Local Similarity 74.1%; Pred. No. 9.2e+03;  
 Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 4 gctgcagcttgaccattggcaccacac 30  
 ||||| ||||| ||||| ||||| |||||  
 Db 56 GTTGCAATATAGACCAATTCCTGCATCA 82

Search completed: June 4, 2000, 13:53:34  
 Job time: 20958 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:09:47 ; Search time 244.64 Seconds  
(without alignments)  
20.191 Million cell updates/sec

Title: US-09-164-714-18

Perfect score: 38  
Sequence: 1 catgctgcagcttgaccatttggcaccacaacataagg 38

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 375880

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/PCFUS.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/1na/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.6	51.6	96	1	US-07-835-029A-2
2	19.6	51.6	96	1	US-08-049-511-2
3	19.6	51.6	96	1	US-07-934-554-2
4	19.6	51.6	96	6	PCT-US93-07877-2
5	19.6	51.6	96	6	PCT-US94-04151-2
6	17.4	45.8	63	6	PCT-US95-13975-63
7	16.6	43.7	45	1	US-08-083-948-4
8	16.6	43.7	45	1	US-08-393-785-4
9	16.6	43.7	45	1	US-08-475-694-4
10	16.6	43.7	45	2	US-08-712-057-4
11	16.6	43.2	60	1	US-07-734-225A-11
12	16.4	43.2	60	2	US-07-692-995B-11
13	16.4	43.2	60	2	US-08-488-457-11
14	16.4	43.2	60	3	US-08-338-793D-11
15	16.4	43.2	60	3	US-08-431-459A-11
16	16.4	43.2	90	6	PCT-US95-13975-70
17	16.4	42.1	39	1	US-08-361-920-32
18	16.4	42.1	39	1	US-08-479-939-32
19	16.4	42.1	39	2	US-08-483-432-32
20	15.8	41.6	45	3	US-08-750-856A-4
21	15.4	40.5	50	1	US-08-207-901-75
22	15.4	40.5	63	6	PCT-US95-13975-49
23	15.2	40.0	51	1	US-08-207-996-10
24	15.2	40.0	51	4	US-08-760-840A-10
25	15.2	40.0	63	1	US-07-734-225A-12
26	15.2	40.0	63	1	US-07-692-995B-12
27	15.2	40.0	63	2	US-08-488-457-12

28	15.2	40.0	63	3	US-08-338-793D-12	Sequence 12, Appl
29	15.2	40.0	63	3	US-08-431-459A-12	Sequence 12, Appl
30	15.2	40.0	80	2	US-07-971-160-50	Sequence 50, Appl
31	15.2	40.0	80	1	US-08-336-241-50	Sequence 50, Appl
32	15.2	40.0	80	3	US-08-465-273-50	Sequence 50, Appl
33	15.2	40.0	80	4	US-09-119-024-50	Sequence 50, Appl
34	15.2	40.0	80	4	US-08-417-226-50	Sequence 50, Appl
35	15.2	40.0	85	1	US-07-741-931-5	Sequence 5, Appl1
36	15.2	40.0	85	1	US-07-937-132A-5	Sequence 5, Appl1
37	15.2	40.0	85	1	US-07-937-132A-6	Sequence 6, Appl1
38	15.2	40.0	86	1	US-07-741-931-7	Sequence 7, Appl1
39	15.2	40.0	86	1	US-07-937-132A-7	Sequence 7, Appl1
40	15.2	39.5	34	4	US-08-459-135A-2	Sequence 2, Appl1
41	14.8	38.9	98	3	US-08-731-272A-4	Sequence 4, Appl1
42	14.4	37.9	36	1	US-07-961-884A-21	Sequence 21, Appl
43	14.4	37.9	36	2	US-08-428-941-21	Sequence 21, Appl
44	14.4	37.9	63	6	PCT-US95-13975-15	Sequence 15, Appl
45	14.4	37.9	63	6	PCT-US95-13975-42	Sequence 42, Appl

## ALIGNMENTS

RESULT 1  
US-07-835-029A-2/C  
Sequence 2, Application US/07835029A  
Patent No. 5284664  
GENERAL INFORMATION:  
APPLICANT: WAGLE, SUDHAKAR S  
APPLICANT: STEINBACH, THOMAS  
APPLICANT: LAWYER, CARL H  
APPLICANT: HERMANN, WILLIAM J  
APPLICANT: GAWISH, ALI ABDEL SALAM  
TITLE OF INVENTION: Method of Treating Presenile or Senile  
TITLE OF INVENTION: Dementia  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Tilton, Fallon, Lungmus & Chestnut  
STREET: 100 South Wacker Drive, Hartford Plaza  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606-4002  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/835,029A  
CLASSIFICATION: 424  
FILING DATE: 19920205  
ATTORNEY/AGENT INFORMATION:  
NAME: Fentress, Susan B  
REGISTRATION NUMBER: 31,327  
REFERENCE/DOCKET NUMBER: 91179A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/456-8000  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 96 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-07-835-029A-2

Query Match 51.6%; Score 19.6; DB 1; Length 96;  
Best Local Similarity 73.5%; Pred. No. 11;  
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 2 atgctgcagcttgaccatttggcaccacaata 35

Db 79 AAGCTGACGCTTGACGTATTATTGACCTTGATA 46

## RESULT 2

US-08-049-511-2/c  
; Sequence 2, Application US/08049511  
; Patent No. 5316775  
; GENERAL INFORMATION:  
; APPLICANT: WAGLE, SUDHAKAR S  
; APPLICANT: STEINBACH, THOMAS  
; APPLICANT: LAWYER, CARL H  
; APPLICANT: HERMANN, WILLIAM J  
; APPLICANT: GAMISH, ALI ABDEL SALAM  
; TITLE OF INVENTION: METHOD OF TREATING  
; TITLE OF INVENTION: HEPATITIS B INFECTION  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TILTON, FALLON, LUNGUMUS & CHESTNUT  
; STREET: 100 SOUTH WACKER DRIVE, SUITE 960  
; CITY: CHICAGO  
; STATE: ILLINOIS  
; COUNTRY: USA  
; ZIP: 60606-4002  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/049,511  
; FILING DATE: 19930419  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/780,084  
; FILING DATE: 15-OCT-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/804,844  
; FILING DATE: 04-DEC-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/728,267  
; FILING DATE: 11-JUL-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/228,364  
; FILING DATE: 04-AUG-1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FENTRESS, SUSAN B  
; REGISTRATION NUMBER: 31,327  
; REFERENCE/DOCKET NUMBER: 92008A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/456-8000  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 96 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
US-08-049-511-2

Query Match 51.6%; Score 19.6; DB 1; Length 96;  
Best Local Similarity 73.5%; Pred. No. 11;  
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 atgtcgcagcttgaccatttggcaccacaata 35  
| | | | | | | | | | | | | | | | | | | | | |  
Db 79 AAGCTGACGCTTGACGTATTATTGACCTTGATA 46

## RESULT 3

US-07-934-554-2/c  
; Sequence 2, Application US/07934554

; Patent No. 5334395  
; GENERAL INFORMATION:  
; APPLICANT: WAGLE, SUDHAKAR S  
; APPLICANT: STEINBACH, THOMAS  
; APPLICANT: LAWYER, CARL H  
; APPLICANT: HERMANN, WILLIAM J  
; APPLICANT: GAMISH, ALI ABDEL SALAM  
; TITLE OF INVENTION: METHOD OF TREATING EPSTEIN-BARR VIRUS INFECTION  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Tilton, Fallon, Lungmus & Chestnut  
; STREET: 100 South Wacker Drive, Hartford Plaza  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606-4002  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/934,554  
; FILING DATE: 19920824  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fentress, Susan B  
; REGISTRATION NUMBER: 31,327  
; REFERENCE/DOCKET NUMBER: 92011A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/456-8000  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 96 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
US-07-934-554-2

Query Match 51.6%; Score 19.6; DB 1; Length 96;  
Best Local Similarity 73.5%; Pred. No. 11;  
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 atgtcgcagcttgaccatttggcaccacaata 35  
| | | | | | | | | | | | | | | | | | | | | |  
Db 79 AAGCTGACGCTTGACGTATTATTGACCTTGATA 46

## RESULT 4

PCT-US93-07877-2/c  
; Sequence 2, Application PC/TUS9307877  
; GENERAL INFORMATION:  
; APPLICANT: WAGLE, SUDHAKAR S  
; APPLICANT: STEINBACH, THOMAS  
; APPLICANT: LAWYER, CARL H  
; APPLICANT: HERMANN, WILLIAM J  
; APPLICANT: GAMISH, ALI ABDEL SALAM  
; TITLE OF INVENTION: METHOD OF TREATING EPSTEIN-BARR VIRUS INFECTION  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Tilton, Fallon, Lungmus & Chestnut  
; STREET: 100 South Wacker Drive, Hartford Plaza  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606-4002  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07877
; FILING DATE: 19930823
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Feitress, Susan B
; REGISTRATION NUMBER: 31,327
; REFERENCE/DOCKET NUMBER: 92011A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/456-8000
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; PCT-US93-07877-2

Query Match          51.6%; Score 19.6; DB 6; Length 96;
Best Local Similarity 73.5%; Pred. No. 11;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 atgtcgacgttgaccatttggccacacata 35
Db 79 AAGCTGCAGCTTGACGTATTATTGCACTTGGA 46

RESULT 5
PCT-US94-04151-2/c
; Sequence 2, Application PC/TUS9404151
; GENERAL INFORMATION:
; APPLICANT: WAGLE, SUDHAKAR S
; APPLICANT: STEINBACH, THOMAS
; APPLICANT: LAWYER, CARL H
; APPLICANT: HERMANN, WILLIAM J
; APPLICANT: GAMISH, ALI ABDEL SALAM
; TITLE OF INVENTION: METHOD OF TREATING HEPATITIS B
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TILTON, FALLON, LUNGWUS & CHESTNUT
; STREET: 100 SOUTH WACKER DRIVE, SUITE 960
; CITY: CHICAGO
; STATE: ILLINOIS
; COUNTRY: USA
; ZIP: 60606-4002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04151
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/7780,084
; FILING DATE: 15-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/804,844
; FILING DATE: 04-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/728,267
; FILING DATE: 11-JUL-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/228,364
; FILING DATE: 04-AUG-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: FEITRESS, SUSAN B
; REGISTRATION NUMBER: 31,327
; REFERENCE/DOCKET NUMBER: 92008A
; PCT-US93-07877-2
```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/456-8000
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; PCT-US94-04151-2

Query Match          51.6%; Score 19.6; DB 6; Length 96;
Best Local Similarity 73.5%; Pred. No. 11;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 atgtcgacgttgaccatttggccacacata 35
Db 79 AAGCTGCAGCTTGACGTATTATTGCACTTGGA 46

RESULT 6
PCT-US95-13975-63
; Sequence 63, Application PC/TUS9513975
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: McDONELL, Michael W.
; TITLE OF INVENTION: Recombinant feline Herpes virus
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM 330 466 DX2
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13975
; FILING DATE: 26-OCT-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/329,883
; FILING DATE: 26-OCT-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 39118-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)278-0400
; TELEFAX: (212)391-0525
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 63 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Pseudorabies virus; Feline herpesvirus
; PCT-US95-13975-63

Query Match          45.8%; Score 17.4; DB 6; Length 63;
Best Local Similarity 77.8%; Pred. No. 79;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

Qy 4 gctgcagcttgaccatttgcacca 30  
||||| | | | | | | | | |  
Db 36 GCTGCAGGTCGACGAGTCTACACCA 62

## RESULT 7

US-08-083-948-4  
; Sequence 4, Application US/08083948  
; Patent No. 5518908  
; GENERAL INFORMATION:  
; APPLICANT: Corbin, David R.  
; APPLICANT: Greenplate, John T.  
; APPLICANT: Jennings, Michael G.  
; APPLICANT: Purcell, John P.  
; APPLICANT: Sammons, Robert D.  
; TITLE OF INVENTION: Method of Controlling Insects  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Grace L. Bonner, Monsanto Company, BB4F  
; STREET: 700 Chesterfield Parkway No. 5518908th  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63198  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/083,948  
; FILING DATE: 19930628  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/762682  
; FILING DATE: 23-SEP-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937195  
; FILING DATE: 09-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bonner, Grace L.  
; REGISTRATION NUMBER: 32,963  
; REFERENCE/DOCKET NUMBER: 38-21(10631)A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (314)537-7286  
; TELEFAX: (314)537-6047  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 45 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-083-948-4

Query Match 43.7%; Score 16.6; DB 1; Length 45;  
Best Local Similarity 71.0%; Pred. No. 1.5e+02;  
Matches 22; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 2 atgctgcagcttgaccatttgcaccaac 32  
||||| | | | | | | | | |  
Db 13 ATGCTGAGATGGCCAGCTGTGAACACG 43

## RESULT 8

US-08-393-785-4  
; Sequence 4, Application US/08393785  
; Patent No. 5554369  
; GENERAL INFORMATION:  
; APPLICANT: Corbin, David R.  
; APPLICANT: Greenplate, John T.

APPLICANT: Jennings, Michael G.  
APPLICANT: Purcell, John P.  
APPLICANT: Sammons, Robert D.  
TITLE OF INVENTION: Method of Controlling Insects  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Grace L. Bonner, Monsanto Company, BB4F  
STREET: 700 Chesterfield Parkway No. 5554369th  
CITY: St. Louis  
STATE: Missouri  
COUNTRY: USA  
ZIP: 63198

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/393,785  
FILING DATE: 24-FEB-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/083,948  
FILING DATE: 28-JUN-1993  
APPLICATION NUMBER: US 07/762682  
FILING DATE: 23-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937195  
FILING DATE: 09-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Bonner, Grace L.  
REGISTRATION NUMBER: 32,963  
REFERENCE/DOCKET NUMBER: 38-21(10631)A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (314)537-7286  
TELEFAX: (314)537-6047  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 45 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-393-785-4

Query Match 43.7%; Score 16.6; DB 1; Length 45;  
Best Local Similarity 71.0%; Pred. No. 1.5e+02;  
Matches 22; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 2 atgctgcagcttgaccatttgcaccaac 32  
||||| | | | | | | | | |  
Db 13 ATGCTGAGATGGCCAGCTGTGAACACG 43

## RESULT 9

US-08-475-694-4  
; Sequence 4, Application US/08475694  
; Patent No. 5558862  
; GENERAL INFORMATION:  
; APPLICANT: Corbin, David R.  
; APPLICANT: Greenplate, John T.  
; APPLICANT: Jennings, Michael G.  
; APPLICANT: Purcell, John P.  
; APPLICANT: Sammons, Robert D.  
; TITLE OF INVENTION: Method of Controlling Insects  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Grace L. Bonner, Monsanto Company, BB4F  
; STREET: 700 Chesterfield Parkway No. 5558862th  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA



```

;
; ZIP: 63198
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,694
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/393,785
; FILING DATE: 24-FEB-1995
; APPLICATION NUMBER: US 08/083,948
; FILING DATE: 28-JUN-1993
; APPLICATION NUMBER: US 07/762682
; FILING DATE: 23-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937195
; FILING DATE: 09-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Bonner, Grace L.
; REGISTRATION NUMBER: 32,963
; REFERENCE/DOCKET NUMBER: 38-21(10631)A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314)537-7286
; TELEFAX: (314)537-6047
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-475-694-4

Query Match      43.7%; Score 16.6; DB 1; Length 45;
Best Local Similarity 71.0%; Pred. No. 1.5e+02;
Matches 22; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY      2 atgtcgagcttgaccatttggcaccac 32
        ||||| | | | | | | | | | | | |
DB      13 ATGCTGGAGATGGCCAGCTGTGGAACACG 43

RESULT 10
US-08-712-057-4
; Sequence 4, Application US/08712057
; Patent No. 5763245
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Greenplate, John T.
; APPLICANT: Purcell, John P.
; APPLICANT: Pershing, Jay
; TITLE OF INVENTION: Method of Controlling Insects
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Grace L. Bonner, Monsanto Company, B94F
; STREET: 700 Chesterfield Parkway No. 5763245th
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63198
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/712,057
; FILING DATE:
; CLASSIFICATION: 800
```

```

;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/762682
; FILING DATE: 23-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937195
; FILING DATE: 09-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Bonner, Grace L.
; REGISTRATION NUMBER: 32,963
; REFERENCE/DOCKET NUMBER: 38-21(10631)A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314)537-7286
; TELEFAX: (314)537-6047
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-712-057-4

Query Match      43.7%; Score 16.6; DB 2; Length 45;
Best Local Similarity 71.0%; Pred. No. 1.5e+02;
Matches 22; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY      2 atgtcgagcttgaccatttggcaccac 32
        ||||| | | | | | | | | | | | |
DB      13 ATGCTGGAGATGGCCAGCTGTGGAACACG 43

RESULT 11
US-07-734-225A-11/C
; Sequence 11, Application US/07734225A
; Patent No. 5320840
; GENERAL INFORMATION:
; APPLICANT: Camble, Roger
; APPLICANT: Timms, David
; APPLICANT: Wilkinson, Anthony J.
; TITLE OF INVENTION: CONTINUOUS RELEASE PHARMACEUTICAL
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: Eleventh Floor, 1615 L. Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/734,225A
; FILING DATE: 19910722
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9016138.1
; FILING DATE: 23-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9018414.4
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9018415.1
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9018416.9
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9018417.1
```

```
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9018418.5
; FILING DATE: 23-AUG-1990
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)861-3000
; TELEFAX: (202)822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 60 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-07-734-225A-11

Query Match          43.2%; Score 16.4; DB 1; Length 60;
Best Local Similarity 67.6%; Pred. No. 2e+02;
Matches 23; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 gctgcagcttgacccaatttgaccacacatag 37
    |||||  || |||  |||||  |||
Db 34 GCTGCAGTGTGTCACAGGTGGCCCAATTCAGG 1

RESULT 12
US-07-692-995B-11/c
; Sequence 11, Application US/07692995B
; Patent No. 5416195
; GENERAL INFORMATION:
; APPLICANT: Camble, Roger
; APPLICANT: Heather, Carr
; APPLICANT: Timms, David
; APPLICANT: Wilkinson, Anthony J.
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: Eleventh Floor, 1615 L. Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/692,995B
; FILING DATE: 19910429
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9009623.1
; FILING DATE: 30-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9013773.8
; FILING DATE: 20-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9016215.7
; FILING DATE: 24-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9102799.5
; FILING DATE: 11-FEB-1991
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 248453 CUSH
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 60 base pairs
; TYPE: NUCLEIC ACID
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-07-692-995B-11

Query Match          43.2%; Score 16.4; DB 1; Length 60;
Best Local Similarity 67.6%; Pred. No. 2e+02;
Matches 23; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 gctgcagcttgacccaatttgaccacacatag 37
    |||||  || |||  |||||  |||
Db 34 GCTGCAGTGTGTCACAGGTGGCCCAATTCAGG 1

RESULT 13
US-08-488-457-11/c
; Sequence 11, Application US/08488457
; Patent No. 5773581
; GENERAL INFORMATION:
; APPLICANT: Camble, Roger
; APPLICANT: Timms, David
; APPLICANT: Wilkinson, Anthony J.
; TITLE OF INVENTION: CONTINUOUS RELEASE PHARMACEUTICAL
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: Eleventh Floor, 1615 L. Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,457
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/155,327
; FILING DATE: 22-NOV-1993
; APPLICATION NUMBER: US 07/734,225
; FILING DATE: 22-JUL-1991
; APPLICATION NUMBER: GB 9016138.1
; FILING DATE: 23-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9018414.4
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9018415.1
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9018416.9
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9018417.1
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9018418.5
; FILING DATE: 23-AUG-1990
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)861-3000
; TELEFAX: (202)822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 60 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```



---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:05:44 ; Search time 1236.38 Seconds  
(without alignments)  
-31.472 Million cell updates/sec

Title: US-09-164-714-19  
Perfect score: 40  
Sequence: 1 cactctgcagtagcagccaaagcaatcacggtagaaca 40

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 356616

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pr1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vl:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_or:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vl:\*  
32: gb\_htg1:\*  
33: gb\_htg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pr4:\*  
41: gb\_htg3:\*  
42: gb\_htg4:\*  
43: gb\_htg5:\*  
44: gb\_htg6:\*

45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pr5:\*  
52: gb\_htg8:\*  
53: gb\_htg9:\*  
54: gb\_htg10:\*  
55: gb\_htg11:\*  
56: gb\_htg12:\*  
57: gb\_htg13:\*  
58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	19.2	48.0	54	5	A57959 Sequence 25
2	18.2	45.5	54	5	A57952 Sequence 18
3	18.2	45.5	54	5	A57960 Sequence 26
4	17.2	43.0	54	5	A57944 Sequence 10
5	17.2	43.0	54	5	A57945 Sequence 11
6	17.2	43.0	54	5	A57946 Sequence 12
7	17.2	43.0	54	5	A57947 Sequence 13
8	17.2	43.0	54	5	A57948 Sequence 14
9	17.2	43.0	54	5	A57949 Sequence 15
10	17.2	43.0	54	5	A57951 Sequence 17
11	17.2	43.0	54	5	A57954 Sequence 20
12	17.2	43.0	54	5	A57956 Sequence 22
13	17.2	43.0	54	5	A57957 Sequence 23
14	17.2	43.0	54	5	A57961 Sequence 27
15	17.2	43.0	54	5	E05823 DNA encodin
16	16.8	42.0	54	5	A57950 Sequence 16
17	16.8	42.0	54	5	A57955 Sequence 21
18	16.8	42.0	84	16	HCVHR31 299037 Hepatitis C
19	16.8	42.0	84	16	HCVHR33 299039 Hepatitis C
20	16.8	42.0	93	9	D78279S17 D78295 Homo sapien
21	16.2	40.5	54	5	A57958 Sequence 24
22	16.2	40.5	57	5	I70194 Sequence 9
23	16.2	40.0	69	5	AR035210 Sequence
24	16.2	40.0	69	5	AR035215 Sequence
25	15.8	39.5	35	5	A21055 Synthetic O
26	15.8	39.5	74	12	RATPAM25 U52661 Rattus norv
27	15.8	39.5	96	9	HS001820 A0001820 Homo sapi
28	15.8	39.5	97	5	I91489 Sequence 23
29	15.6	39.0	54	5	A57943 Sequence 9
30	15.6	39.0	54	5	A57953 Sequence 19
31	15.6	39.0	54	5	A57965 Sequence 31
32	15.6	39.0	58	5	E01587 DNA sequenc
33	15.6	39.0	58	5	E01745 DNA sequenc
34	15.6	39.0	61	5	E01585 DNA sequenc
35	15.6	39.0	84	16	HCVHR10 299016 Hepatitis C
36	15.6	39.0	84	16	HCVHR11 299017 Hepatitis C
37	15.6	39.0	84	16	HCVHR9 299015 Hepatitis C
38	15.4	38.5	65	1	S72520 flm3-major
39	15.4	38.5	65	5	A21052 B.pertussis
40	15.2	38.0	81	12	MMR25B9 X53627 M.musculus
41	15.2	38.0	84	16	HCVHR12 299018 Hepatitis C
42	15.2	38.0	84	16	HCVHR13 299019 Hepatitis C
43	15.2	38.0	84	16	HCVHR15 299021 Hepatitis C
44	15.2	38.0	84	16	HCVHR16 299022 Hepatitis C
45	15.2	38.0	84	16	HCVHR17 299023 Hepatitis C

ALIGNMENTS

```
RESULT 1
LOCUS A57959 54 bp DNA
DEFINITION Sequence 25 from Patent EP0743364.
ACCESSION A57959
VERSION A57959.1 GI:3713729
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS 1 (bases 1 to 54)
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding
JOURNAL fragments and their application as reactives for risk evaluation of
COMMENT Patent: EP 0743364-A 25-20-NOV-1996;
FEATURES COMMISSARIAT ENERGIE ATOMIQUE (FR)
SOURCE Location/Qualifiers
BASE COUNT 30 a 4 c 15 g 5 t
ORIGIN

Query Match
Best Local Similarity 48.0%; Score 19.2; DB 5; Length 54;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 9 agtagacgcgaagaatcaacgctaagaaca 40
Db 18 AGTAGAGGAAGACCAAAACAAAGTAGAAGAAA 49

RESULT 2
LOCUS A57952 54 bp DNA
DEFINITION Sequence 18 from Patent EP0743364.
ACCESSION A57952
VERSION A57952.1 GI:3713722
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS 1 (bases 1 to 54)
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding
JOURNAL fragments and their application as reactives for risk evaluation of
COMMENT Patent: EP 0743364-A 18-20-NOV-1996;
FEATURES COMMISSARIAT ENERGIE ATOMIQUE (FR)
SOURCE Location/Qualifiers
BASE COUNT 29 a 5 c 15 g 4 t 1 others
ORIGIN

Query Match
Best Local Similarity 45.5%; Score 18.2; DB 5; Length 54;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 10 gtagacgcgaagaatcaacgctaagaaca 40
Db 19 GTAGAGGAAGACCAAAACAAAGTAGAAGAAA 49

RESULT 3
LOCUS A57960 54 bp DNA
DEFINITION Sequence 26 from Patent EP0743364.
ACCESSION A57960
VERSION A57960.1 GI:3713730
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS 1 (bases 1 to 54)
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding
JOURNAL fragments and their application as reactives for risk evaluation of
COMMENT Patent: EP 0743364-A 26-20-NOV-1996;
FEATURES COMMISSARIAT ENERGIE ATOMIQUE (FR)
SOURCE Location/Qualifiers
BASE COUNT 30 a 4 c 15 g 5 t
ORIGIN

Query Match
Best Local Similarity 45.5%; Score 18.2; DB 5; Length 54;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 10 gtagacgcgaagaatcaacgctaagaaca 40
Db 19 GTAGAGGAAGACCAAAACAAAGTAGAAGAAA 49

RESULT 4
LOCUS A57944 54 bp DNA
DEFINITION Sequence 10 from Patent EP0743364.
ACCESSION A57944
VERSION A57944.1 GI:3713714
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS 1 (bases 1 to 54)
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding
JOURNAL fragments and their application as reactives for risk evaluation of
COMMENT Patent: EP 0743364-A 10-20-NOV-1996;
FEATURES COMMISSARIAT ENERGIE ATOMIQUE (FR)
SOURCE Location/Qualifiers
BASE COUNT 31 a 4 c 13 g 6 t
ORIGIN

Query Match
Best Local Similarity 43.0%; Score 17.2; DB 5; Length 54;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 11 tagacgcgaagaatcaacgctaagaaca 40
Db 20 TAGAGGAAGACCAAAACAAAGTAGAAGAAA 49

RESULT 5
LOCUS A57945 54 bp DNA
DEFINITION Sequence 11 from Patent EP0743364.
ACCESSION A57945
```

```
LOCUS A57960 54 bp DNA
DEFINITION Sequence 26 from Patent EP0743364.
ACCESSION A57960
VERSION A57960.1 GI:3713730
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS 1 (bases 1 to 54)
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding
JOURNAL fragments and their application as reactives for risk evaluation of
COMMENT Patent: EP 0743364-A 26-20-NOV-1996;
FEATURES COMMISSARIAT ENERGIE ATOMIQUE (FR)
SOURCE Location/Qualifiers
BASE COUNT 30 a 4 c 15 g 5 t
ORIGIN

Query Match
Best Local Similarity 45.5%; Score 18.2; DB 5; Length 54;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 10 gtagacgcgaagaatcaacgctaagaaca 40
Db 19 GTAGAGGAAGACCAAAACAAAGTAGAAGAAA 49

RESULT 4
LOCUS A57944 54 bp DNA
DEFINITION Sequence 10 from Patent EP0743364.
ACCESSION A57944
VERSION A57944.1 GI:3713714
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS 1 (bases 1 to 54)
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding
JOURNAL fragments and their application as reactives for risk evaluation of
COMMENT Patent: EP 0743364-A 10-20-NOV-1996;
FEATURES COMMISSARIAT ENERGIE ATOMIQUE (FR)
SOURCE Location/Qualifiers
BASE COUNT 31 a 4 c 13 g 6 t
ORIGIN

Query Match
Best Local Similarity 43.0%; Score 17.2; DB 5; Length 54;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 11 tagacgcgaagaatcaacgctaagaaca 40
Db 20 TAGAGGAAGACCAAAACAAAGTAGAAGAAA 49

RESULT 5
LOCUS A57945 54 bp DNA
DEFINITION Sequence 11 from Patent EP0743364.
ACCESSION A57945
```

VERSION A57945.1 GI:3713715  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 11 20-NOV-1996;  
COMMISSARIAT ENERGIE ATOMIQUE (FR)  
COMMENT Other publication FR 2734281 961122.  
FEATURES  
source 1..54  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 30 a 4 c 15 g 5 t  
ORIGIN

Query Match 43.0%; Score 17.2; DB 5; Length 54;  
Best Local Similarity 73.3%; Pred. No. 6.4e+03;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 11 tagagccaagaacaatcaacggtagaaca 40  
||||| ||||| ||| ||||| |||  
Db 20 TAGAGGAGAGCAAAACAAAGTAAGAAAA 49

RESULT 6  
LOCUS A57946 54 bp DNA PAT 05-MAR-1998  
DEFINITION Sequence 12 from Patent EP0743364.  
ACCESSION A57946  
VERSION A57946.1 GI:3713716  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 12 20-NOV-1996;  
COMMISSARIAT ENERGIE ATOMIQUE (FR)  
COMMENT Other publication FR 2734281 961122.  
FEATURES  
source 1..54  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 31 a 4 c 14 g 5 t  
ORIGIN

Query Match 43.0%; Score 17.2; DB 5; Length 54;  
Best Local Similarity 73.3%; Pred. No. 6.4e+03;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 11 tagagccaagaacaatcaacggtagaaca 40  
||||| ||||| ||| ||||| |||  
Db 20 TAGAGGAGAGCAAAACAAAGTAAGAAAA 49

RESULT 7  
LOCUS A57947 54 bp DNA PAT 05-MAR-1998  
DEFINITION Sequence 13 from Patent EP0743364.  
ACCESSION A57947  
VERSION A57947.1 GI:3713717  
KEYWORDS  
SOURCE  
unidentified.

ORGANISM  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 13 20-NOV-1996;  
COMMISSARIAT ENERGIE ATOMIQUE (FR)  
COMMENT Other publication FR 2734281 961122.  
FEATURES  
source 1..54  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 31 a 4 c 13 g 6 t  
ORIGIN

Query Match 43.0%; Score 17.2; DB 5; Length 54;  
Best Local Similarity 73.3%; Pred. No. 6.4e+03;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 11 tagagccaagaacaatcaacggtagaaca 40  
||||| ||||| ||| ||||| |||  
Db 20 TAGAGGAGAGCAAAACAAAGTAAGAAAA 49

RESULT 8  
LOCUS A57948 54 bp DNA PAT 05-MAR-1998  
DEFINITION Sequence 14 from Patent EP0743364.  
ACCESSION A57948  
VERSION A57948.1 GI:3713718  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 14 20-NOV-1996;  
COMMISSARIAT ENERGIE ATOMIQUE (FR)  
COMMENT Other publication FR 2734281 961122.  
FEATURES  
source 1..54  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 32 a 5 c 12 g 5 t  
ORIGIN

Query Match 43.0%; Score 17.2; DB 5; Length 54;  
Best Local Similarity 73.3%; Pred. No. 6.4e+03;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 11 tagagccaagaacaatcaacggtagaaca 40  
||||| ||||| ||| ||||| |||  
Db 20 TAGAGGAGAGCAAAACAAAGTAAGAAAA 49

RESULT 9  
LOCUS A57949 54 bp DNA PAT 05-MAR-1998  
DEFINITION Sequence 15 from Patent EP0743364.  
ACCESSION A57949  
VERSION A57949.1 GI:3713719  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 54)

AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 15-20-NOV-1996;  
COMMENT COMMISSARIAT ENERGIE ATOMIQUE (FR)  
FEATURES Other publication FR 2734281 961122.  
source Location/Qualifiers  
1. 54  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 30 a 5 c 14 g 5 t  
ORIGIN

Query Match 43.0%; Score 17.2; DB 5; Length 54;  
Best Local Similarity 73.3%; Pred. No. 6.4e+03;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 11 tagagccaagcaatcaacgtaagaaca 40  
||||| ||||||| ||| ||||||| |  
Db 20 TAGAGGAAGCAGCAAAACAAAGTAGAGAAA 49

RESULT 10  
A57951 54 bp DNA PAT 05-MAR-1998  
LOCUS A57951  
DEFINITION Sequence 17 from Patent EP0743364.  
ACCESSION A57951  
VERSION A57951.1 GI:3713721  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 17-20-NOV-1996;  
COMMENT COMMISSARIAT ENERGIE ATOMIQUE (FR)  
FEATURES Other publication FR 2734281 961122.  
source Location/Qualifiers  
1. 54  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 31 a 4 c 14 g 5 t  
ORIGIN

Query Match 43.0%; Score 17.2; DB 5; Length 54;  
Best Local Similarity 73.3%; Pred. No. 6.4e+03;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 11 tagagccaagcaatcaacgtaagaaca 40  
||||| ||||||| ||| ||||||| |  
Db 20 TAGAGGAAGCAGCAAAACAAAGTAGAGAAA 49

RESULT 11  
A57954 54 bp DNA PAT 05-MAR-1998  
LOCUS A57954  
DEFINITION Sequence 20 from Patent EP0743364.  
ACCESSION A57954  
VERSION A57954.1 GI:3713724  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of

JOURNAL HIV-1 mother-foetal transmission  
COMMENT Patent: EP 0743364-A 20-20-NOV-1996;  
FEATURES COMMISSARIAT ENERGIE ATOMIQUE (FR)  
source Other publication FR 2734281 961122.  
Location/Qualifiers  
1. 54  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 30 a 4 c 15 g 5 t  
ORIGIN

Query Match 43.0%; Score 17.2; DB 5; Length 54;  
Best Local Similarity 73.3%; Pred. No. 6.4e+03;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 11 tagagccaagcaatcaacgtaagaaca 40  
||||| ||||||| ||| ||||||| |  
Db 20 TAGAGGAAGCAGCAAAACAAAGTAGAGAAA 49

RESULT 12  
A57956 54 bp DNA PAT 05-MAR-1998  
LOCUS A57956  
DEFINITION Sequence 22 from Patent EP0743364.  
ACCESSION A57956  
VERSION A57956.1 GI:3713726  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 22-20-NOV-1996;  
COMMENT COMMISSARIAT ENERGIE ATOMIQUE (FR)  
FEATURES Other publication FR 2734281 961122.  
source Location/Qualifiers  
1. 54  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 30 a 6 c 14 g 4 t  
ORIGIN

Query Match 43.0%; Score 17.2; DB 5; Length 54;  
Best Local Similarity 73.3%; Pred. No. 6.4e+03;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 11 tagagccaagcaatcaacgtaagaaca 40  
||||| ||||||| ||| ||||||| |  
Db 20 TAGAGGAAGCAGCAAAACAAAGTAGAGAAA 49

RESULT 13  
A57957 54 bp DNA PAT 05-MAR-1998  
LOCUS A57957  
DEFINITION Sequence 23 from Patent EP0743364.  
ACCESSION A57957  
VERSION A57957.1 GI:3713727  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa,R. and Roques,P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-foetal transmission  
JOURNAL Patent: EP 0743364-A 23-20-NOV-1996;  
COMMENT COMMISSARIAT ENERGIE ATOMIQUE (FR)



COMMENT Other publication FR 2734281 961122.  
FEATURES Location/Qualifiers  
SOURCE 1..54  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 30 a 6 c 13 g 5 t  
ORIGIN

Query Match 43.0%; Score 17.2; DB 5; Length 54;  
Best Local Similarity 73.3%; Pred. No. 6.4e+03;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 11 tagaccgccaagccaatcacaacggtagaaca 40  
||||| ||||| ||| ||||| |||  
Db 20 TAGAGGAGAGACCAACCAAGTAAGAAAA 49

RESULT 14  
LOCUS A57961 54 bp DNA PAT 05-MAR-1998  
DEFINITION Sequence 27 from Patent EP0743364.  
ACCESSION A57961  
VERSION A57961.1 GI:3713731  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 54)  
AUTHORS Narwa, R. and Roques, P.  
TITLE Nucleic acid fragments derived from the HIV-1 genome, corresponding fragments and their application as reactives for risk evaluation of HIV-1 mother-to-fetal transmission  
JOURNAL Patent: EP 0743364-A 27 20-NOV-1996;  
COMMENT COMMISSARIAT ENERGIE ATOMIQUE (FR)  
FEATURES Other publication FR 2734281 961122.  
SOURCE 1..54  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 30 a 4 c 15 g 5 t  
ORIGIN

Query Match 43.0%; Score 17.2; DB 5; Length 54;  
Best Local Similarity 73.3%; Pred. No. 6.4e+03;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 11 tagaccgccaagccaatcacaacggtagaaca 40  
||||| ||||| ||| ||||| |||  
Db 20 TAGAGGAGAGACCAACCAAGTAAGAAAA 49

RESULT 15  
LOCUS E05823 92 bp DNA PAT 29-SEP-1997  
DEFINITION DNA encoding HGP-30 of HIV.  
ACCESSION E05823  
VERSION E05823.1 GI:2174010  
KEYWORDS JP 1993301895-A/1.  
SOURCE JP 1993301895-A/1.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 92)  
AUTHORS Nagaya, A., Takamura, C. and Kamogawa, K.  
TITLE HYBRID ANTIGEN PROTEIN, RECOMBINED VIRUS FOR EXPRESSING THE SAME, AND ITS PRODUCTION  
JOURNAL Patent: JP 1993301895-A 1 16-NOV-1993;  
COMMENT NIPPON ZEON CO LTD  
OS Artificial gene  
OC Artificial sequence; Genes.  
PN JP 1993301895-A/1  
PD 16-NOV-1993

PF 22-APR-1992 JP 1992127980  
PI NAGAYA ANSUSHI, TAKAMURA CHIZUKO, KAMOGAWA KOICHI PC  
C07K15/04,A61K39/00,A61K39/00,A61K39/275,C12N7/01, PC  
C12P21/02//C12N15/39,  
PC C12N15/62,C12N15/86,(C12P21/02,C12R1:92);  
CC strandedness: Double;  
CC topology: Linear;  
FEATURES Location/Qualifiers  
SOURCE 1..92  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 47 a 10 c 22 g 13 t  
ORIGIN

Query Match 43.0%; Score 17.2; DB 5; Length 92;  
Best Local Similarity 73.3%; Pred. No. 6.5e+03;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 11 tagaccgccaagccaatcacaacggtagaaca 40  
||||| ||||| ||| ||||| |||  
Db 55 TAGAGGAGAGACCAACCAAGTAAGAAAA 84

Search completed: June 4, 2000, 16:05:44  
Job time: 27896 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:24:21 ; Search time 322.35 Seconds  
(without alignments)  
31.046 Million cell updates/sec

Title: US-09-164-714-19

Perfect score: 40  
Sequence: 1 cactctgcagtagacgccaagcaatcaacgtaagaaca 40

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 431286

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.2	48.0	54	1 T43657	HIV-1 matrix prote
2	18.2	45.5	54	1 T43658	HIV-1 matrix prote
3	18.2	45.5	54	1 T43650	HIV-1 matrix prote
4	17.2	43.0	54	1 T43659	HIV-1 matrix prote
5	17.2	43.0	54	1 T43642	HIV-1 matrix prote
6	17.2	43.0	54	1 T43643	HIV-1 matrix prote
7	17.2	43.0	54	1 T43644	HIV-1 matrix prote
8	17.2	43.0	54	1 T43645	HIV-1 matrix prote
9	17.2	43.0	54	1 T43646	HIV-1 matrix prote
10	17.2	43.0	54	1 T43647	HIV-1 matrix prote
11	17.2	43.0	54	1 T43649	HIV-1 matrix prote
12	17.2	43.0	54	1 T43652	HIV-1 matrix prote
13	17.2	43.0	54	1 T43654	HIV-1 matrix prote
14	17.2	43.0	54	1 T43655	HIV-1 matrix prote
15	17.2	43.0	92	1 05311	Hybrid antigen pro
16	16.8	42.0	54	1 T43648	HIV-1 matrix prote
17	16.8	42.0	54	1 T43653	HIV-1 matrix prote
18	16.4	41.0	35	1 V29846	Human CD4 transmem
19	16.2	40.5	54	1 T43656	HIV-1 matrix prote
20	16.2	40.5	57	1 T28274	Primer lux-5'-BamH
21	16	40.0	69	1 030897	Primer 312-69. New
22	16	40.0	69	1 030892	Primer 312-64. New
23	16	40.0	96	1 T49233	HVRI region of E2
24	15.8	39.5	97	1 T19416	Human gene signalu
25	15.6	39.0	46	1 T79510	Primer Strep-tag o
26	15.6	39.0	47	1 X52594	Human genome blatt
27	15.6	39.0	47	1 X52596	Human genome blatt
28	15.6	39.0	47	1 X52546	Human genome blatt
29	15.6	39.0	47	1 X52544	Human genome blatt
30	15.6	39.0	54	1 T43653	Human genome blatt
31	15.6	39.0	54	1 T43651	HIV-1 matrix prote
32	15.2	38.0	38	1 T43641	HIV-1 matrix prote
33	15.2	38.0	38	1 N94509	Probe for N-termin
34	15.2	38.0	38	1 V83523	PCR primer used to

PCR primer used to  
Oligonucleotide us  
Quadruplex/duplex  
Fibroblast growth  
Fibroblast growth  
Primer 13 for 95 k  
PCR primer used to  
Staphylococcus aur  
DNA sequence encod  
Modified type of s  
DNA sequence of 1m

## ALIGNMENTS

RESULT 1  
T43657  
ID T43657 standard; DNA: 54 BP.  
AC T43657;  
DT 19-AUG-1997 (first entry)  
DE HIV-1 matrix protein p17 gene fragment 4541.  
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
KW detection; maternal transmission; hybridisation assay; immunoassay; ss.  
OS Human immunodeficiency virus type 1.  
PN Ep-743364-A2.  
PD 20-NOV-1996.  
PE 17-MAY-1996; 401084.  
PR 18-MAY-1995; FR-005914.  
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
PI Narwa R, Rogues P;  
DR WP1: 96-507733/51.  
PT Human immunodeficiency virus p17 gene fragments, derived proteins  
PT and antibodies - useful for assessing the risk of maternal  
PT transmission of HIV-1 infection  
PS Claim 3: Page 27: 46pp: French.  
CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, CTA, GTC or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-fetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
CC Sequence 54 BP; 30 A; 4 C; 15 G; 5 T;  
SO

Query Match 48.0%; Score 19.2; DB 1; Length 54;  
Best local Similarity 75.0%; Pred. No. 27;  
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 9 agtagcgcagcaacatcaacgtaagaaca 40  
||||| | ||||| ||| ||||| |  
Db 18 AGTAGAGGAGAGCAACAAACAAAGTAGAGAA 49

RESULT 2  
T43658  
ID T43658 standard; DNA: 54 BP.  
AC T43658;  
DT 19-AUG-1997 (first entry)  
DE HIV-1 matrix protein p17 gene fragment 2754.  
KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
KW detection; maternal transmission; hybridisation assay; immunoassay; ss.  
OS Human immunodeficiency virus type 1.  
PN Ep-743364-A2.  
PD 20-NOV-1996.  
PE 17-MAY-1996; 401084.

PR 18-MAY-1995; FR-005914.  
 PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
 PI Narwa R. Roques P;  
 DR WPI: 96-507733/51.  
 PT Human immunodeficiency virus p17 gene fragments, derived proteins  
 PT and antibodies - useful for assessing the risk of maternal  
 PT transmission of HIV-1 infection  
 PS Claim 3; Page 27; 46pp; French.  
 CC This sequence is a specifically claimed example of 21-90 nucleotide  
 CC long nucleic acid fragments, derived from the gene encoding part of the  
 CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
 CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
 CC Y6 = ATG, TTA, CTC, GTA, CTA, CTG or ATG; Y7 = GAG or GAA; Y8 = GAA or  
 CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
 CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
 CC = AAA or AAG, then Y6 is not ATG, Y7 is not GAG or GAA, Y8 is not GAA,  
 CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
 CC The new nucleic acid fragments can be used as reagents for determining  
 CC and assessing the risk of maternal-fetal transmission of HIV-1, using  
 CC standard hybridisation or immuno assays. The presence of such sequences  
 CC in maternal blood is strongly correlated with transmission of infection.  
 SQ Sequence 54 BP; 30 A; 4 C; 15 G; 5 T;

Query Match 45.5%; Score 18.2; DB 1; Length 54;  
 Best Local Similarity 74.2%; Pred. No. 67;  
 Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 10 gtagagccaagaatcaacgtaagaaca 40  
 ||||| | ||||| ||| ||||| |||  
 Db 19 GTAGAGCAAGAGCAAAACAAAGTAAGAAAA 49

RESULT 3  
 T43650 standard; DNA; 54 BP.  
 AC T43650;  
 DE 19-AUG-1997 (first entry)  
 DT HIV-1 matrix protein p17 gene fragment HAR.  
 KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
 KM detection; maternal transmission; hybridisation assay; immunoassay; ss.  
 OS Human immunodeficiency virus type 1.  
 FT key location/Qualifiers  
 FT mat\_peptide 1..54  
 FT /\*tag= a  
 FT /product= HAR\_peptide

EP-743364-A2.  
 PD 20-NOV-1996.  
 PR 17-MAY-1996; 401084.  
 PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
 PI Narwa R. Roques P;  
 DR WPI: 96-507733/51.  
 P-PSDB: W06612.  
 PT Human immunodeficiency virus p17 gene fragments, derived proteins  
 PT and antibodies - useful for assessing the risk of maternal  
 PT transmission of HIV-1 infection  
 PS Claim 3; Page 25; 46pp; French.  
 CC This sequence is a specifically claimed example of 21-90 nucleotide  
 CC long nucleic acid fragments, derived from the gene encoding part of the  
 CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
 CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
 CC Y6 = ATG, TTA, CTC, GTA, CTA, CTG or ATG; Y7 = GAG or GAA; Y8 = GAA or  
 CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
 CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
 CC = AAA or AAG, then Y6 is not ATG, Y7 is not GAG or GAA, Y8 is not GAA,  
 CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
 CC The new nucleic acid fragments can be used as reagents for determining  
 CC and assessing the risk of maternal-fetal transmission of HIV-1, using  
 CC standard hybridisation or immuno assays. The presence of such sequences  
 CC in maternal blood is strongly correlated with transmission of infection.  
 SQ Sequence 54 BP; 29 A; 5 C; 15 G; 4 T;

Query Match 45.5%; Score 18.2; DB 1; Length 54;  
 Best Local Similarity 74.2%; Pred. No. 67;  
 Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 10 gtagagccaagaatcaacgtaagaaca 40  
 ||||| | ||||| ||| ||||| |||  
 Db 19 GTAGAGCAAGAGCAAAACAAAGTAAGAAAA 49

RESULT 4  
 T43659 standard; DNA; 54 BP.  
 AC T43659;  
 DE 19-AUG-1997 (first entry)  
 DT HIV-1 matrix protein p17 gene fragment 2826.  
 KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
 KM detection; maternal transmission; hybridisation assay; immunoassay; ss.  
 OS Human immunodeficiency virus type 1.  
 FT EP-743364-A2.  
 PD 20-NOV-1996.  
 PR 17-MAY-1996; 401084.  
 PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
 PI Narwa R. Roques P;  
 DR WPI: 96-507733/51.  
 PT Human immunodeficiency virus p17 gene fragments, derived proteins  
 PT and antibodies - useful for assessing the risk of maternal  
 PT transmission of HIV-1 infection  
 PS Claim 3; Page 27; 46pp; French.  
 CC This sequence is a specifically claimed example of 21-90 nucleotide  
 CC long nucleic acid fragments, derived from the gene encoding part of the  
 CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
 CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
 CC Y6 = ATG, TTA, CTC, GTA, CTA, CTG or ATG; Y7 = GAG or GAA; Y8 = GAA or  
 CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
 CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
 CC = AAA or AAG, then Y6 is not ATG, Y7 is not GAG or GAA, Y8 is not GAA,  
 CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
 CC The new nucleic acid fragments can be used as reagents for determining  
 CC and assessing the risk of maternal-fetal transmission of HIV-1, using  
 CC standard hybridisation or immuno assays. The presence of such sequences  
 CC in maternal blood is strongly correlated with transmission of infection.  
 SQ Sequence 54 BP; 30 A; 4 C; 15 G; 5 T;

Query Match 43.0%; Score 17.2; DB 1; Length 54;  
 Best Local Similarity 73.3%; Pred. No. 1.7e+02;  
 Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 11 tagagccaagaatcaacgtaagaaca 40  
 |||| | ||||| ||| ||||| |||  
 Db 20 TAGAGCAAGAGCAAAACAAAGTAAGAAAA 49

RESULT 5  
 T43642 standard; DNA; 54 BP.  
 AC T43642;  
 DE 19-AUG-1997 (first entry)  
 DT HIV-1 matrix protein p17 gene fragment ARI.  
 KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
 KM detection; maternal transmission; hybridisation assay; immunoassay; ss.  
 OS Human immunodeficiency virus type 1.  
 FT EP-743364-A2.  
 PD 20-NOV-1996.  
 PR 17-MAY-1996; 401084.  
 PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
 PI Narwa R. Roques P;  
 DR WPI: 96-507733/51.  
 PT Human immunodeficiency virus p17 gene fragments, derived proteins  
 PT and antibodies - useful for assessing the risk of maternal

PT transmission of HIV-1 infection.  
 CS Claim 3: Page 23: 46pp: French.  
 CC This sequence is a specifically claimed example of 21-90 nucleotide  
 CC long nucleic acid fragments, derived from the gene encoding part of the  
 CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
 CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
 CC Y6 = ATA, TTA, CTG, CTA, GTC or ATG; Y7 = GAG or GAA; Y8 = GAA or  
 CC CAG; Y9 = ATA, GTA, CTA, GTC, GAC, GAT or AGA; Y10 = CAA, CAG or  
 CC CCA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
 CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
 CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
 CC The new nucleic acid fragments can be used as reagents for determining  
 CC and assessing the risk of maternal-fetal transmission of HIV-1, using  
 CC standard hybridisation or immuno assays. The presence of such sequences  
 CC in maternal blood is strongly correlated with transmission of infection.  
 SQ Sequence 54 BP: 31 A; 4 C; 13 G; 6 T;  
 Query Match 43.0%; Score 17.2; DB 1; Length 54;  
 Best Local Similarity 73.3%; Pred. No. 1.7e+02;  
 Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 11 tagacgccaagcaatacaagcgtagaaca 40  
 Db 20 TAGAGGAAGAGCAAAACAAAGTAGAGAAA 49  
 RESULT 6  
 ID T43643 standard; DNA; 54 BP.  
 AC T43643:  
 DT 19-AUG-1997 (first entry)  
 DE HIV-1 matrix protein p17 gene fragment BOI.  
 KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
 KM detection; maternal transmission; hybridisation assay; immunoassay; ss.  
 OS Human immunodeficiency virus type 1.  
 FH Key location/Qualifiers  
 FT mat\_peptide 1..54  
 FT /\*tag- a /product= BOI\_peptide  
 FT PN EP-743364-A2.  
 FT PD 20-NOV-1996.  
 FT PF 17-MAY-1995; FR-005914.  
 FT PR 18-MAY-1995; FR-005914.  
 FT PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
 FT PI Narwa R., Roques P;  
 FT DR WPI: 96-507733/51.  
 FT DR P-PSDB: W06610.  
 PT Human immunodeficiency virus p17 gene fragments, derived proteins  
 PT and antibodies - useful for assessing the risk of maternal  
 PT transmission of HIV-1 infection  
 PS Claim 3: Page 23: 46pp: French.  
 CC This sequence is a specifically claimed example of 21-90 nucleotide  
 CC long nucleic acid fragments, derived from the gene encoding part of the  
 CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
 CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
 CC Y6 = ATA, TTA, CTG, CTA, GTC or ATG; Y7 = GAG or GAA; Y8 = GAA or  
 CC CAG; Y9 = ATA, GTA, CTA, GTC, GAC, GAT or AGA; Y10 = CAA, CAG or  
 CC CCA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
 CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
 CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
 CC The new nucleic acid fragments can be used as reagents for determining  
 CC and assessing the risk of maternal-fetal transmission of HIV-1, using  
 CC standard hybridisation or immuno assays. The presence of such sequences  
 CC in maternal blood is strongly correlated with transmission of infection.  
 SQ Sequence 54 BP: 30 A; 4 C; 15 G; 5 T;  
 Query Match 43.0%; Score 17.2; DB 1; Length 54;  
 Best Local Similarity 73.3%; Pred. No. 1.7e+02;  
 Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 11 tagacgccaagcaatacaagcgtagaaca 40

Db 20 TAGAGGAAGAGCAAAACAAAGTAGAGAAA 49  
 RESULT 7  
 ID T43644 standard; DNA; 54 BP.  
 AC T43644:  
 DT 19-AUG-1997 (first entry)  
 DE HIV-1 matrix protein p17 gene fragment DUM.  
 KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
 KM detection; maternal transmission; hybridisation assay; immunoassay; ss.  
 OS Human immunodeficiency virus type 1.  
 PN EP-743364-A2.  
 PD 20-NOV-1996.  
 PF 17-MAY-1995; FR-005914.  
 PF 18-MAY-1995; FR-005914.  
 PF PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
 PF PI Narwa R., Roques P;  
 PF DR WPI: 96-507733/51.  
 PT Human immunodeficiency virus p17 gene fragments, derived proteins  
 PT and antibodies - useful for assessing the risk of maternal  
 PT transmission of HIV-1 infection  
 PS Claim 3: Page 24: 46pp: French.  
 CC This sequence is a specifically claimed example of 21-90 nucleotide  
 CC long nucleic acid fragments, derived from the gene encoding part of the  
 CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
 CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
 CC Y6 = ATA, TTA, CTG, CTA, GTC or ATG; Y7 = GAG or GAA; Y8 = GAA or  
 CC CAG; Y9 = ATA, GTA, CTA, GTC, GAC, GAT or AGA; Y10 = CAA, CAG or  
 CC CCA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
 CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
 CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
 CC The new nucleic acid fragments can be used as reagents for determining  
 CC and assessing the risk of maternal-fetal transmission of HIV-1, using  
 CC standard hybridisation or immuno assays. The presence of such sequences  
 CC in maternal blood is strongly correlated with transmission of infection.  
 SQ Sequence 54 BP: 31 A; 4 C; 14 G; 5 T;  
 Query Match 43.0%; Score 17.2; DB 1; Length 54;  
 Best Local Similarity 73.3%; Pred. No. 1.7e+02;  
 Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 11 tagacgccaagcaatacaagcgtagaaca 40  
 Db 20 TAGAGGAAGAGCAAAACAAAGTAGAGAAA 49  
 RESULT 8  
 ID T43645 standard; DNA; 54 BP.  
 AC T43645:  
 DT 19-AUG-1997 (first entry)  
 DE HIV-1 matrix protein p17 gene fragment PAL.  
 KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
 KM detection; maternal transmission; hybridisation assay; immunoassay; ss.  
 OS Human immunodeficiency virus type 1.  
 PN EP-743364-A2.  
 PD 20-NOV-1996.  
 PF 17-MAY-1995; FR-005914.  
 PF 18-MAY-1995; FR-005914.  
 PF PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
 PF PI Narwa R., Roques P;  
 PF DR WPI: 96-507733/51.  
 PT Human immunodeficiency virus p17 gene fragments, derived proteins  
 PT and antibodies - useful for assessing the risk of maternal  
 PT transmission of HIV-1 infection  
 PS Claim 3: Page 24: 46pp: French.  
 CC This sequence is a specifically claimed example of 21-90 nucleotide  
 CC long nucleic acid fragments, derived from the gene encoding part of the  
 CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
 CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;

CC Y6 = ATA, TTA, CTG, GTA, CTA, GTG or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-foetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 31 A; 4 C; 13 G; 6 T;

Query Match 43.0%; Score 17.2; DB 1; Length 54;  
Best Local Similarity 73.3%; Pred. No. 1.7e+02;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 11 tagagcgcaagcaaatcaacgtaagaaca 40  
Db 20 TAGAGGAGAGCAAAACAAAGTAGAGAAA 49

RESULT 9

ID T43646 standard; DNA; 54 BP.

AC T43646;

DT 19-AUG-1997 (first entry)

DE HIV-1 matrix protein p17 gene fragment RYO.

KM Human immunodeficiency virus: matrix protein p17; prognosis; probe;

OS detection; maternal transmission; hybridisation assay; immunoassay; ss.

PN Human immunodeficiency virus type 1.

PD BP-743364-A2.

PP 20-NOV-1996.

PR 17-MAY-1996; 401084.

PR 18-MAY-1995; FR-005914.

PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.

PI Narva R. Roques P.

DR WPI; 96-507733/51.

PT Human immunodeficiency virus p17 gene fragments, derived proteins

PT and antibodies - useful for assessing the risk of maternal

PT transmission of HIV-1 infection

PS Claim 3: Page 24; 46pp: French.

CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, GTA, CTA, GTG or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-foetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 32 A; 5 C; 12 G; 5 T;

Query Match 43.0%; Score 17.2; DB 1; Length 54;

Best Local Similarity 73.3%; Pred. No. 1.7e+02;

Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 11 tagagcgcaagcaaatcaacgtaagaaca 40  
Db 20 TAGAGGAGAGCAAAACAAAGTAGAGAAA 49

RESULT 10

ID T43647 standard; DNA; 54 BP.

AC T43647;

DT 19-AUG-1997 (first entry)

DE HIV-1 matrix protein p17 gene fragment FLO.

KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;

KM detection; maternal transmission; hybridisation assay; immunoassay; ss.

OS Human immunodeficiency virus type 1.

PN BP-743364-A2.

PP 20-NOV-1996.

PR 17-MAY-1996; 401084.

PR 18-MAY-1995; FR-005914.

PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.

PI Narva R. Roques P.

DR WPI; 96-507733/51.

PT Human immunodeficiency virus p17 gene fragments, derived proteins

PT and antibodies - useful for assessing the risk of maternal

PT transmission of HIV-1 infection

PS Claim 3: Page 24; 46pp: French.

CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, GTA, CTA, GTG or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-foetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 30 A; 5 C; 14 G; 5 T;

Query Match 43.0%; Score 17.2; DB 1; Length 54;

Best Local Similarity 73.3%; Pred. No. 1.7e+02;

Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 11 tagagcgcaagcaaatcaacgtaagaaca 40  
Db 20 TAGAGGAGAGCAAAACAAAGTAGAGAAA 49

RESULT 11

ID T43649 standard; DNA; 54 BP.

AC T43649;

DT 19-AUG-1997 (first entry)

DE HIV-1 matrix protein p17 gene fragment FAL.

KM Human immunodeficiency virus: matrix protein p17; prognosis; probe;

OS detection; maternal transmission; hybridisation assay; immunoassay; ss.

PN Human immunodeficiency virus type 1.

PD BP-743364-A2.

PP 20-NOV-1996.

PR 17-MAY-1996; 401084.

PR 18-MAY-1995; FR-005914.

PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.

PI Narva R. Roques P.

DR WPI; 96-507733/51.

PT Human immunodeficiency virus p17 gene fragments, derived proteins

PT and antibodies - useful for assessing the risk of maternal

PT transmission of HIV-1 infection

PS Claim 3: Page 25; 46pp: French.

CC This sequence is a specifically claimed example of 21-90 nucleotide  
CC long nucleic acid fragments, derived from the gene encoding part of the  
CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
CC Y6 = ATA, TTA, CTG, GTA, CTA, GTG or ATG; Y7 = GAG or GAA; Y8 = GAA or  
CC AAG; Y9 = ATA, GTA, CTA, GAA, GAG, GAC, GAT or AGA; Y10 = CAA, CAG or  
CC CGA; Y11 = AAA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
CC The new nucleic acid fragments can be used as reagents for determining  
CC and assessing the risk of maternal-foetal transmission of HIV-1, using  
CC standard hybridisation or immuno assays. The presence of such sequences  
CC in maternal blood is strongly correlated with transmission of infection.  
SQ Sequence 54 BP; 31 A; 4 C; 14 G; 5 T;

Query Match 43.0%; Score 17.2; DB 1; Length 54;  
 Best Local Similarity 73.3%; Pred. No. 1.7e+02;  
 Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 11 tagagcgcaagcaatcaacgtaagaaca 40  
 ||||| ||||| ||||| ||||| ||||| ||  
 Db 20 TAGAGGAGAGCAACAAACAAAGTAGAGAAA 49

RESULT 12  
 T43652 standard; DNA; 54 BP.  
 ID T43652  
 AC T43652;  
 DT 19-AUG-1997 (first entry)  
 DE HIV-1 matrix protein p17 gene fragment CHET.  
 KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
 KM detection; maternal transmission; hybridisation assay; immunoassay; ss.  
 OS Human immunodeficiency virus type 1.  
 PN EP-743364-A2.  
 PD 20-NOV-1996.  
 PF 17-MAY-1996; 401084.  
 PR 18-MAY-1995; FR-005914.  
 PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
 PI Narwa R. Roques P;  
 PI WPI: 96-507733/51.  
 DR Human immunodeficiency virus p17 gene fragments, derived proteins  
 PT and antibodies - useful for assessing the risk of maternal  
 PS transmission of HIV-1 infection  
 PT Claim 3; Page 26; 46pp; French.  
 CC This sequence is a specifically claimed example of 21-90 nucleotide  
 CC long nucleic acid fragments, derived from the gene encoding part of the  
 CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
 CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
 CC Y6 = ATA, TTA, CTG, GTA, CTA, GAG, GAC, GAT or AGA; Y7 = GAG or GAA; Y8 = GAA or  
 CC AAG; Y9 = ATA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
 CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
 CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
 CC The new nucleic acid fragments can be used as reagents for determining  
 CC and assessing the risk of maternal-foetal transmission of HIV-1, using  
 CC standard hybridisation or immuno assays. The presence of such sequences  
 CC in maternal blood is strongly correlated with transmission of infection.  
 CC Sequence 54 BP; 30 A; 4 C; 15 G; 5 T;

Query Match 43.0%; Score 17.2; DB 1; Length 54;  
 Best Local Similarity 73.3%; Pred. No. 1.7e+02;  
 Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 11 tagagcgcaagcaatcaacgtaagaaca 40  
 ||||| ||||| ||||| ||||| ||||| ||  
 Db 20 TAGAGGAGAGCAACAAACAAAGTAGAGAAA 49

RESULT 13  
 T43654 standard; DNA; 54 BP.  
 ID T43654;  
 AC T43654;  
 DT 19-AUG-1997 (first entry)  
 DE HIV-1 matrix protein p17 gene fragment STM.  
 KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
 KM detection; maternal transmission; hybridisation assay; immunoassay; ss.  
 OS Human immunodeficiency virus type 1.  
 FH Key Location/Qualifiers  
 FT mat\_peptide 1..54  
 FT /tag- a  
 FT /product- STM\_peptide  
 PN EP-743364-A2.  
 PD 20-NOV-1996.  
 PF 17-MAY-1996; 401084.  
 PR 18-MAY-1995; FR-005914.  
 PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.

PI Narwa R. Roques P;  
 DR WPI: 96-507733/51.  
 DR P-PSDB: W06614.  
 PT Human immunodeficiency virus p17 gene fragments, derived proteins  
 PT and antibodies - useful for assessing the risk of maternal  
 PS transmission of HIV-1 infection  
 PS Claim 3; Page 26; 46pp; French.  
 CC This sequence is a specifically claimed example of 21-90 nucleotide  
 CC long nucleic acid fragments, derived from the gene encoding part of the  
 CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
 CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
 CC Y6 = ATA, TTA, CTG, GTA, CTA, GAG, GAC, GAT or AGA; Y7 = GAG or GAA; Y8 = GAA or  
 CC AAG; Y9 = ATA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
 CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
 CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
 CC The new nucleic acid fragments can be used as reagents for determining  
 CC and assessing the risk of maternal-foetal transmission of HIV-1, using  
 CC standard hybridisation or immuno assays. The presence of such sequences  
 CC in maternal blood is strongly correlated with transmission of infection.  
 CC Sequence 54 BP; 30 A; 6 C; 14 G; 4 T;

Query Match 43.0%; Score 17.2; DB 1; Length 54;  
 Best Local Similarity 73.3%; Pred. No. 1.7e+02;  
 Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 11 tagagcgcaagcaatcaacgtaagaaca 40  
 ||||| ||||| ||||| ||||| ||||| ||  
 Db 20 TAGAGGAGAGCAACAAACAAAGTAGAGAAA 49

RESULT 14  
 T43655 standard; DNA; 54 BP.  
 ID T43655  
 AC T43655;  
 DT 19-AUG-1997 (first entry)  
 DE HIV-1 matrix protein p17 gene fragment MOE.  
 KW Human immunodeficiency virus; matrix protein p17; prognosis; probe;  
 KM detection; maternal transmission; hybridisation assay; immunoassay; ss.  
 OS Human immunodeficiency virus type 1.  
 FH Key Location/Qualifiers  
 FT mat\_peptide 1..54  
 FT /tag- a  
 FT /product- MOE\_peptide  
 PN EP-743364-A2.  
 PD 20-NOV-1996.  
 PF 17-MAY-1996; 401084.  
 PR 18-MAY-1995; FR-005914.  
 PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
 PI Narwa R. Roques P;  
 PI WPI: 96-507733/51.  
 DR P-PSDB: W06615.  
 PT Human immunodeficiency virus p17 gene fragments, derived proteins  
 PT and antibodies - useful for assessing the risk of maternal  
 PS transmission of HIV-1 infection  
 PS Claim 3; Page 26; 46pp; French.  
 CC This sequence is a specifically claimed example of 21-90 nucleotide  
 CC long nucleic acid fragments, derived from the gene encoding part of the  
 CC p17 HIV-1 matrix protein; the claimed nucleic acids must include at  
 CC least the sequence: Y5-Y6-Y7-Y8-Y9-Y10-Y11, where Y5 = AAA, AAG or GAA;  
 CC Y6 = ATA, TTA, CTG, GTA, CTA, GAG, GAC, GAT or AGA; Y7 = GAG or GAA; Y8 = GAA or  
 CC AAG; Y9 = ATA, AGT, AAG, AAT, CAT, AAC or ACC; provided that when Y5  
 CC = AAA or AAG, then Y6 is not ATA, Y7 is not GAG or GAA, Y8 is not GAA,  
 CC Y9 is not GAG or GAA, Y10 is not CAA or CAG and Y11 is not AAT or AAC.  
 CC The new nucleic acid fragments can be used as reagents for determining  
 CC and assessing the risk of maternal-foetal transmission of HIV-1, using  
 CC standard hybridisation or immuno assays. The presence of such sequences  
 CC in maternal blood is strongly correlated with transmission of infection.  
 CC Sequence 54 BP; 30 A; 6 C; 13 G; 5 T;

Query Match 43.0%; Score 17.2; DB 1; Length 54;  
Best Local Similarity 73.3%; Pred. No. 1.7e+02;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 11 tagagccaagcaatcaacgtaagaaca 40  
||||| | ||||| ||| ||||| |  
DB 20 TAGAGGAGAGCAACCAAGTAAGAAAA 49

## RESULT 15

053311  
ID 053311 standard; DNA; 92 BP.  
AC 053311.  
DT 09-JUN-1994 (first entry)  
DE Hybrid antigen protein linker sequence #1.  
KW linker; recombinant virus; hybrid antigen protein; virus;  
structural protein; peptide epitope; antigen; vaccine; ss.  
OS Synthetic.  
PN J05301895-A.  
PD 16-NOV-1993.  
PF 22-APR-1992; 127980.  
PR 22-APR-1992; JP-127980.  
PA (JAPG ) NIPPON ZEON KK.  
DR WPI: 93-400397/50.  
PT New hybrid antigen protein and recombinant virus, useful for  
active vaccine - comprises virus structural protein and peptide  
epitope contg. below 300 amino acid(s) useful for antigen of  
PT component vaccine and expressed by recombinant virus  
PS Disclosure: Page 7; 14pp; Japanese.  
CC The sequences given in 053311-12 are linkers which were used in the  
production of a recombinant virus encoding a hybrid antigen protein.  
CC The hybrid antigen protein has a virus structural protein and a peptide  
epitope and consists of less than 300 amino acids. This hybrid  
CC antigen protein may be useful as the antigen component in a vaccine.  
SQ Sequence 92 BP; 47 A; 10 C; 22 G; 13 T;

Query Match 43.0%; Score 17.2; DB 1; Length 92;  
Best Local Similarity 73.3%; Pred. No. 1.8e+02;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 11 tagagccaagcaatcaacgtaagaaca 40  
||||| | ||||| ||| ||||| |  
DB 55 TAGAGGAGAGCAACCAAGTAAGAAAA 84

Search completed: June 4, 2000, 16:24:21  
Job time: 28917 sec



Gencore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 13:53:34 ; Search time 4521.53 Seconds  
(without alignments)  
35.857 Million cell updates/sec

Title: US-09-164-714-19

Perfect score: 40  
Sequence: 1 cactctgcagtacgacccaagcaatacaacgtaagaaca 40

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 119968

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result

Query

No.	Score	Match	Length	DB	ID	Description
1	19	47.5	52	48	AI584790	fb83d08.y
2	18.6	46.5	94	31	AA291480	zt40b04.s
3	17.6	44.0	74	43	AI224006	qx02a05.x
4	17.2	43.0	96	21	DA5646	HMMG502842
5	16.6	41.5	43	43	AI198732	qf78c08.x
6	16.6	41.5	67	44	AI286661	ub95902.r
7	16.4	41.0	93	37	AA713894	nv79b04.s
8	16.2	40.5	93	28	AA066146	mma42d04.r
9	15.8	39.5	89	20	D11978	HMM0512E03
10	15.8	39.5	94	20	D12000	HMM0513A08
11	15.6	39.0	76	61	AI872794	wm70f12.x
12	15.6	39.0	79	39	AA868541	ak43e11.s
13	15.6	39.0	84	42	AF027883	AF027883
14	15.4	38.5	64	37	AA715443	nv53b08.r
15	15.4	38.5	70	49	AI659613	tu06b03.x
16	15.4	38.5	84	83	AF088162	Hom0.sapi
17	15.4	38.5	89	44	AI252064	qv39f11.x
18	15.4	38.5	98	29	AA152894	mq56f02.r
19	15.4	38.5	99	80	AA285906	LG1.251.G
20	15.4	38.5	100	22	R76016	v122f09.r1
21	15.4	38.5	100	37	AA680523	T3670.Blo
22	15.4	38.5	100	44	AI252291	qv26f04.x
23	15.2	38.0	78	49	F24150	HSPD10213.H
24	15.2	38.0	82	82	FR0020751	F.rubripe
25	15.2	38.0	100	63	AI954642	wq34e02.x
26	15	37.5	34	43	AI218040	qh29a03.x
27	15	37.5	59	63	AI988246	sc98a02.y
28	15	37.5	64	27	AA041720	mj05b07.r
29	15	37.5	64	27	W99959	mg29d03.r1
30	15	37.5	67	30	AA218242	mv74d04.r
31	15	37.5	73	48	AI569601	to44c05.x
32	15	37.5	76	37	AA711088	vt55e05.x
33	15	37.5	78	34	AA512518	v118b04.r
34	15	37.5	84	41	AI053592	q168h09.x
35	15	37.5	85	50	AI688982	cx91f08.x
36	15	37.5	100	41	AI061242	an36c05.x
37	14.8	37.0	65	44	AI271266	qw82g05.x
38	14.8	37.0	91	25	N84555	J0022P.Huma
39	14.8	37.0	93	74	AA192657	x148d03.x
40	14.8	37.0	97	62	AI918993	tu16b08.x
41	14.8	37.0	100	34	AA515148	ng68c05.s
42	14.6	36.5	58	82	HSMC37C10	H. sapiens
43	14.6	36.5	63	38	AA746400	nm62c07.s
44	14.6	36.5	63	84	B04181	csRL-28a10
45	14.6	36.5	64	27	AA052614	mc82a09.r

## ALIGNMENTS

```

RESULT 1
LOCUS AI584790 52 bp mRNA EST 06-APR-1999
DEFINITION fb83d08.y1 zebrafish washu MPING EST Danto rerio cDNA 5' similar to
SEQUENCE.
SM:FRP_MOUSE Q62356 FOLISTATIN-RELATED PROTEIN PRECURSOR ;, mRNA
sequence.
ACCESSION AI584790
VERSION AI584790.1 GI:4570687
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danto rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
Cyprinidae; Cyprinidae; Rasboreinae; Danto.
1 (bases 1 to 52)
Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M.,
Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,F.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.

```

```

TITLE Washu Zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT On May 18, 1998 this sequence version replaced gi:3138646.
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watsn.wustl.edu
cDNA Library Preparation: Matthew Clark, cDNA Library Arrayed by:
Matthew Clark, DNA Sequencing by: Washington University Genome
Sequencing Center Clone distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
RessourcenzentrumUmrImmundatenbank, Berlin, Germany (web address:
www.rzpd.de)
Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand
Seq primer: T3 ET from Amersham
High quality sequence stop: 1.
FEATURES
source location/Qualifiers
1..52
/organism="Danto rerio"
/db_xref="taxon:7955"
/clone_lib="Zebrafish Washu MPING EST"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="XLI-blue MRF"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(OT)15 primer
[5']PACTAGTCTTAGATCGCGCGCGCCCTTTTCTTTTCTT3';
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
analysis were selected following oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones
were sequenced additional times to assess quality
control."
BASE COUNT 15 a 9 c 16 g 12 t
ORIGIN
Query Match 47.5%; Score 19; DB 48; Length 52;
Best Local Similarity 81.5%; Pred. No. 4,2e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy
7 gcagtagacgcccaagaatcaacaggt 33
||||| ||| ||||| ||||| |||
Db 7 gcagagcctcaaaccaatcaacagct 33
RESULT 2
LOCUS AA291480 94 bp mRNA EST 16-MAY-1997
DEFINITION zt40b04.s1 Soares ovary tumor Nshot Homo sapiens cDNA clone
IMAGE:724783 3', similar to TR:G14016 G14016 CYTOCHROME C OXIDASE
POLYPEPTIDE II ;, mRNA sequence.
ACCESSION AA291480
VERSION AA291480.1 GI:1939475
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```

[illegible]

unknown library type		Std Error: 0.00
Insert Length: 638		
Seq primer: -40UP from Gldco.		
Location/Qualifiers		
1. .74		
/organism="Homo sapiens"		
/db_xref="taxon:9606"		
/clone_image="2000144"		
/clone_lib="NCI_CGAP_Lym12"		
/tissue_type="lymphoma, follicular mixed small and large cell"		
/lab_host="DH10B"		
/note="Organ: lymph node; Vector: pCMV-SPORT6; Site_1: Sail; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.25 kb. Life technologies catalog #: 11547-015"		
BASE COUNT	19 a	17 c 14 g 24 t
ORIGIN		
Query Match 44.0%; Score 17.6; DB 43; Length 74;		
Best Local Similarity 71.9%; Pred. No. 1.7e+03;		
Matches	23; Conservative	0; Mismatches 9; Indels 0; Gaps 0;
OY	8 cagtagacgccaagccaatcaacggtgaagac 39	
	11       11 11 11	
Db	41 CAGTACACTCCAGGCGCATTTAAGTCACAGAAC 10	
RESULT	4	
LOCUS	DA5646	96 bp mRNA EST 20-FEB-1995
DEFINITION	HUMGS02842 Human adult lung 3' directed MbolI cDNA Homo sapiens CDNA	
ACCESSION	DA5646	3', mRNA sequence.
VERSION	DA5646.1	GI:662600
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
AUTHORS	1 (bases 1 to 96)	
TITLE	Itoh,K., Okubo,K., Yosii,Y., Yokouchi,H. and Matsubara,K.	
JOURNAL	An expression profile of active genes in human lung	
MEDLINE	DNA Research 1, 279-287 (1994)	
COMMENT	95236275	
Contact: Kohichi Itoh		
Institute for Molecular and Cellular Biology		
Osaka University		
3-1, Yamadaoka, Suita, Osaka, 565, Japan		
Tel: 06-877-5111 x3910		
Fax: 06-877-1922		
Insert Length: 396 Std Error: 0.00		
High quality sequence stop: 35.		
Location/Qualifiers		
1. .96		
/organism="Homo sapiens"		
/db_xref="taxon:9606"		
/clone_lib="Human adult lung 3' directed MbolI cDNA"		
/note="Adult human lung, 3' directed MbolI"		
BASE COUNT	23 a	36 c 14 g 20 t 3 others
ORIGIN		
Query Match 43.0%; Score 17.2; DB 21; Length 96;		
Best Local Similarity 71.0%; Pred. No. 2.6e+03;		
Matches	22; Conservative	0; Mismatches 9; Indels 0; Gaps 0;
OY	1 cactgtcagtagacgccaagccaatcaacg 31	
Db	63 CTCTCTCATTAATACCACAGCATCTCAANG 93	

RESULT 5  
LOCUS A1198732 43 bp mRNA EST 27-JAN-1999  
DEFINITION q78c08.x1 Soares\_fetal\_lung\_NBHL19W Homo sapiens CDNA clone  
IMAGE:1756142 3' similar to TR:Q13363 Q13363 CTBP. ;, mRNA  
sequence.  
A1198732  
ACCESSION A1198732.1 GI:3751338  
VERSION A1198732.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 43)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
Trace considered overall poor quality  
Insert Length: 1190 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 1.  
Location/Qualifiers  
1. 43  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1756142"  
/clone\_1lb="Soares\_fetal\_lung\_NBHL19W"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: lung; Vector: p773D (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - 0190(dT) primer  
[5'-GTATCCATCTGAGTGGAGCGCCGCAATGTTTTTTTTTTT-3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified p773 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M.Felima Bonaudo. This library was constructed  
from the same fetus as the fetal heart library, Soares  
fetal heart NBHL19W."  
BASE COUNT 7 a 10 c 14 g 12 t  
ORIGIN

Query Match 41.5%; Score 16.6; DB 43; Length 43;  
Best Local Similarity 82.6%; Pred. No. 4e+03;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 actctgcagtagacgccaagca 24  
|||||  
Db 42 ACTGTCAGTAGAGCCATCCAA 20

RESULT 6  
LOCUS A1286661 67 bp mRNA EST 24-NOV-1998  
DEFINITION ub95q02.x1 Soares\_mammary\_gland\_NBMG Mus musculus CDNA clone  
IMAGE:1396274 5' similar to SW:JANA\_DROPS P54364 SEX-REGULATED  
PROTEIN JAMUS-A. ;, mRNA sequence.  
A1286661  
ACCESSION A1286661.1 GI:3926414  
VERSION A1286661.1  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 67)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Dec 5, 1997 this sequence version replaced g1:2662749.  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
MGI:907990  
Trace considered overall poor quality  
Possible reversed clone: similarity on wrong strand  
Seq primer: -28m3 rev2 ET from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers  
1. 67  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1396274"  
/clone\_1lb="Soares\_mammary\_gland\_NBMG"  
/sex="male"  
/tissue\_type="mammary gland"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Organ: mammary gland; Vector: p773D-Pac  
(Pharmacia) with a modified polylinker; Site\_1: Not I;  
Site\_2: Eco RI; 1st strand cDNA was primed with a Not I -  
0190(dT) primer [5'-  
TGTATCCATCTGAGTGGAGCGCCGCAATGTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified p773 vector.  
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
constructed and normalized by Bento Soares and M.Felima  
Bonaudo."  
BASE COUNT 12 a 14 c 18 g 23 t  
ORIGIN

Query Match 41.5%; Score 16.6; DB 44; Length 67;  
Best Local Similarity 71.0%; Pred. No. 4.3e+03;  
Matches 22; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 10 gttagacgccaagcaatcaacgtagaaca 40  
|||||  
Db 46 GAAGACGCATCCGATCAACGCTACATCA 16

RESULT 7  
LOCUS AA713869 93 bp mRNA EST 29-DEC-1997  
DEFINITION nv79b04.s1 NCI-CGAP\_Br4 Homo sapiens CDNA clone IMAGE:1235983  
similar to gb:X57809 IG LAMBDA CHAIN C REGIONS (HUMAN);, mRNA  
sequence.  
AA713869  
ACCESSION AA713869.1 GI:2726143  
VERSION AA713869.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 93)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index			
JOURNAL	Unpublished (1997)			
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1407147. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Ian Kirsch, M.D., Kristina A. Cole, M.D., Ph.D. student, Rodrigo F. Chuaguil, M.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: David B. Krizman, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ULNL at: www-bio.1lnl.gov/bdrp/image/image.html			
FEATURES	Trace considered overall poor quality Seq primer: -40ml3 fwd. RT from Amersham High quality sequence stop: 1. Location/Qualifiers 1..93 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1235983" /clone_lib="NCI-CGAP-B14" /sex="female" /tissue_type="normal ductal tissue" /lab_host="DH10B" /note="Organ: breast; Vector: PAMP10; mRNA made from normal breast ductal tissue. cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."			
BASE COUNT	8 a 25 c 27 g 33 t			
ORIGIN				
Query Match	41.0%; Score 16.4; DB 37; Length 93;			
Best Local Similarity	67.6%; Pred. 0.5.5e+03;			
Matches	23; Conservative 0; Mismatches 11; Indels 0; Gaps 0;			
OY	7	gcagtagagccgaagaatcaaggtgaaca	40	
Db	35	GCAGTGGAGACCACACACCCATCCAAACA	2	
RESULT	8			
LOCUS	AA066146	93 bp	mRNA	EST
DEFINITION	mm2d04.r1 Stratagene mouse melanoma (#937312) Mus musculus cDNA clone IMAGE:524167 5' similar to gb:D10049 Mouse mRNA for mouse melanoma antigen, complete cds (MUSEP);, mRNA sequence.			
ACCESSION	AA066146			
VERSION	AA066146.1	GI:1563134		
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 93) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.			
TITLE	The WashU-HMT Mouse EST Project			
JOURNAL	Unpublished (1996)			
COMMENT	On Apr 14, 1993 this sequence version replaced gi:692360. Contact: Marra M/Mouse EST Project WashU-HMT Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108			

```

Tel: 314 286 1800
Fax: 314 286 1810
Email: mouse@est.watson.wustl.edu
This clone is available royalty-free through LINT; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:318015
Seq primer: -28ml3 rev1 ET from Amer sham
High quality sequence stop: 1.
Location/Qualifiers
1. .93
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:524167"
/clone_lib="Stratagene mouse melanoma (#937312)"
/tissue_type="melanoma"
/dev_stage="M2 cells"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. From M2 cells, a highly metastatic derivative of
the K-1735 (mouse) melanoma. Average insert size: 1.0 kb.
Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTGGCCACGAG
3' -3' adaptor sequence: 5' CTCGACTTCTTTTCTTTTCTT 3'"

BASE COUNT      32 a      16 c      22 g      23 t
ORIGIN

Query Match      40.5%; Score 16.2; DB 28; Length 93;
Best Local Similarity 72.4%; Pred. No. 6.7e+03;
Matches 21; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Oy      5      ctgcagtagacgcgcagcaatcaacggt 33
          ||||| 11 ||| 11 ||| 11
Db      86      CTCGACTAGCTCTTACCTTATGAAGGT 58

RESULT 9
D11978 c      89 bp      mRNA      EST      02-DEC-1992
LOCUS      HMW0512E03 Liver HepG2 cell line. Homo sapiens cDNA clone s12e03,
DEFINITION      mRNA sequence.
ACCESSION      D11978
VERSION      D11978.1 GI:2148809
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 89)
Okubo,K., Hori,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y.,
and Matsubara,K.
large scale cDNA sequencing for analysis of quantitative and
qualitative aspects of gene expression
Nature Genet. 2, 173-179 (1992)
94258199
JOURNAL      CONTACT: Kousaku Okubo, Naohiro Hori, Ryo Matoba, Toshiyuki
MEDLINE      Niiyama, Atsushi Fukushima, Yuiko Kojima & Kenichi Matsubara
COMMENT      Institute for Molecular and Cellular Biology
Osaka University
1-3 Yamada-oka, Suita, Osaka 565, Japan.
FEATURES
Source
1. .89
/organism="Homo sapiens"
/db_xref="GDB:D058180E"
/db_xref="taxon:9606"
/clone="s12e03"
/clone_lib="Liver HepG2 cell line."
/lab_host="E.coli"
/note="3'-directed regional cDNA library. Cleaved by MboI
and transformed into E.coli."
14 a      21 c      22 g      32 t
BASE COUNT
ORIGIN

```

Qy	6	tgcagtagacgccaagcaatcaatcaggttaagaaca	40
Db	79	TACCGGAATGCGCAGCAATCATCACAGGAGAGACA	45
RESULT	10		
LOCUS	D12000/c		
DEFINITION	HUM0S13A08 Liver HepG2 cell line. Homo sapiens cDNA clone s13a08,	EST	02-DEC-1992
ACCESION	D12000		
VERSION	D12000.1	GI:2148823	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	Okubo,K., Horii,N., Matoba,R., Niyama,T., Fukushima,A., Kojima,Y. and Matsubara,K. Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression Nature Genet. 2, 173-179 (1992)		
JOURNAL MEDLINE COMMENT	Contact: Kousaku Okubo, Naohiro Horii, Ryo Matoba, Toshiyuki Niyama, Atsushi Fukushima, Yoko Kojima & Kenichi Matsubara Institute for Molecular and Cellular Biology Osaka University 1-3 Yamada-oka,Suita,Osaka 565,Japan. Location/Qualifiers 1..94 /organism="Homo sapiens" /db_xref="GDB:D058194E" /db_xref="taxon:9606" /clone="s13a08" /clone_lib="Liver HepG2 cell line." /lab_host="E.coli" /note="3'-directed regional cDNA library. Cleaved by MboI /note="3'-transformed into E.coli."		
BASE COUNT	18 a	21 c	21 g
ORIGIN		33 t	1 others
Query Match		39.5%;	Score 15.8; DB 20; Length 94;
Best Local Similarity		65.7%;	Pred. No. 9.7e+03;
Matches	23; Conservative	0;	Mismatches 12; Indels 0; Gaps 0;
Qy	6	tgcagtagacgccaagcaatcaatcaggttaagaaca	40
Db	79	TACCGGAATGCGCAGCAATCATCACAGGAGAGACA	45
RESULT	11		
LOCUS	A1872794/c		
DEFINITION	wm70f12.x1 NCI-CGAP,Ut2 Homo sapiens cDNA clone IMAGE:2441327 3'	EST	30-AUG-1999
ACCESION	A1872794		
VERSION	A1872794.1	GI:5546843	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		

```

JOURNAL
COMMENT
Unpublished (1997)
On Dec 20, 1995 this sequence version replaced g1:1134490.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/bdrrp/image/image.html

Seq primer: -40UP from Glibco
High quality sequence stop: 30.
Location/Qualifiers
1. .76
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2441327"
/clone_1bp="NCI_CGAP_Uc2"
/tissue_type="moderately-differentiated endometrial
adenocarcinoma, 3 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"

BASE COUNT
11 a 22 c 11 g 32 t
ORIGIN

Query Match 39.0%; Score 15.6; DB 61; Length 76;
Best Local Similarity 63.2%; Pred. No. 1.1e+04;
Matches 24; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 2 acctgcagtagcagccaagcaatcaacggtaagaac 39
I I I I I I I I I I I I I I I I I I I I
Db 51 AATTGGAAGTCGAAGAAAGAAAGAAAGCAACCAAGAAC 14

RESULT 12
AA868541/c
LOCUS
DEFINITION
aa868541.91 Soares_testis_NHT Homo sapiens CDNA clone IMAGE:1408748
3' similar to WP:F59F4.2 CE1548 ;, mRNA sequence.
ACCESSION
AA868541
VERSION
AA868541.1 GI:2963986
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Euthera; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 79)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Jan 19, 1998 this sequence version replaced g1:2152084.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/bdrrp/image/image.html

Possible reversed clone: similarity on wrong strand
Insert Length: 1180 Std Error: 0.00

```



## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 70)

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Unpublished (1997)  
On Mar 16, 1998 this sequence version replaced gi:2961802.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/bbrp/image/image.html

## FEATURES

## source

Trace considered overall poor quality  
Seq primer: -40UP from Gibco  
High quality sequence stop: 1.  
Location/Qualifiers

1..70  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2250221"  
/clone\_lib="NCI-CGAP\_Pr28"  
/sex="male"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: p77T3D-Pac (Pharmacia)  
with a modified polylinker: Plasmid DNA from the  
normalized library NCI-CGAP\_Pr22 was prepared, and ss  
circles were made in vitro. Following HAP purification,  
this DNA was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from a pool  
of 5,000 clones made from the same library (clones  
985608-986759, 1101192-1101959, and 1217928-1220615).  
Subtraction by Bento Soares and M. Fatima Bonaldo."

## BASE COUNT

3 a 23 c 14 g 30 t

## ORIGIN

## Query Match

38.5%; Score 15.4; DB 49; Length 70;

Best Local Similarity 66.7%; Pred. No. 1.4e+04;  
Matches 22; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

## Qy

6 tagcagtagacgcgaagcaatcaacgtagaag 38  
||||| ||| | | | ||||| |||  
Db 42 TGCACGACAGACGAGAAAAAGCAGCAGAAAA 10

Search completed: June 4, 2000, 13:53:37  
Job time: 20961 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:09:48 ; Search time 244.64 Seconds  
(without alignments)  
21.253 Million cell updates/sec

Title: US-09-164-714-19

Perfect score: 40

Sequence: 1 cactctgcagtagacgcgaagcaatcaacgctagaaca 40

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 373880

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/PCUS.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/1na/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.2	48.0	54	US-08-649-991-25	Sequence 25, Appl
2	18.2	45.5	54	US-08-649-991-18	Sequence 18, Appl
3	18.2	45.5	54	US-08-649-991-26	Sequence 26, Appl
4	18.2	45.5	54	US-08-649-991-120	Sequence 120, Appl
5	17.2	43.0	54	US-08-649-991-10	Sequence 10, Appl
6	17.2	43.0	54	US-08-649-991-11	Sequence 11, Appl
7	17.2	43.0	54	US-08-649-991-12	Sequence 12, Appl
8	17.2	43.0	54	US-08-649-991-13	Sequence 13, Appl
9	17.2	43.0	54	US-08-649-991-14	Sequence 14, Appl
10	17.2	43.0	54	US-08-649-991-15	Sequence 15, Appl
11	17.2	43.0	54	US-08-649-991-17	Sequence 17, Appl
12	17.2	43.0	54	US-08-649-991-20	Sequence 20, Appl
13	17.2	43.0	54	US-08-649-991-22	Sequence 22, Appl
14	17.2	43.0	54	US-08-649-991-23	Sequence 23, Appl
15	17.2	43.0	54	US-08-649-991-27	Sequence 27, Appl
16	17.2	43.0	54	US-08-649-991-119	Sequence 119, Appl
17	17.2	43.0	54	US-08-649-991-121	Sequence 121, Appl
18	17.2	43.0	54	US-08-649-991-122	Sequence 122, Appl
19	17.2	43.0	54	US-08-649-991-123	Sequence 123, Appl
20	17.2	43.0	54	US-08-649-991-124	Sequence 124, Appl
21	17.2	43.0	54	US-08-649-991-125	Sequence 125, Appl
22	17.2	43.0	54	US-08-649-991-126	Sequence 126, Appl
23	17.2	43.0	54	US-08-649-991-128	Sequence 128, Appl
24	17.2	43.0	54	US-08-649-991-129	Sequence 129, Appl
25	16.8	42.0	54	US-08-213-811-10	Sequence 10, Appl
26	16.8	42.0	54	US-08-649-991-16	Sequence 16, Appl
27	16.8	42.0	54	US-08-649-991-21	Sequence 21, Appl

28	16.2	40.5	54	US-08-649-991-24	Sequence 24, Appl
29	16.2	40.5	57	US-08-316-950-9	Sequence 9, Appl
30	16.2	40.5	57	PCT-US95-12642-9	Sequence 9, Appl
31	16	40.0	69	US-07-916-098A-26	Sequence 26, Appl
32	16	40.0	69	US-07-916-098A-26	Sequence 31, Appl
33	15.8	39.5	35	US-08-213-811-9	Sequence 9, Appl
34	15.8	39.5	57	5514566-17	Patent No. 5514566
35	15.8	39.5	7	US-08-447-172A-23	Sequence 23, Appl
36	15.6	39.0	54	US-08-649-991-9	Sequence 9, Appl
37	15.6	39.0	54	US-08-649-991-19	Sequence 19, Appl
38	15.6	39.0	54	US-08-649-991-31	Sequence 31, Appl
39	15.4	38.5	65	US-08-213-811-6	Sequence 6, Appl
40	15	37.5	35	US-08-811-492-93	Sequence 93, Appl
41	15	37.5	60	PCT-US95-11985A-11	Sequence 11, Appl
42	15	37.5	69	US-08-096-623A-42	Sequence 42, Appl
43	15	37.5	71	US-08-095-726-36	Sequence 36, Appl
44	15	37.5	71	US-08-096-043-33	Sequence 33, Appl
45	15	37.5	71	US-08-093-577-29	Sequence 29, Appl

## ALIGNMENTS

RESULT 1  
Sequence 25, Application US/08649991  
Patent No. 5919462  
GENERAL INFORMATION:  
APPLICANT: Narwa, Remy  
TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE  
TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR  
TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP  
STREET: 1800 M Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036-5869  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/649,991  
FILING DATE: 17-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9505914  
FILING DATE: 18-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Adler, Reid G.  
REGISTRATION NUMBER: 30,988  
REFERENCE/DOCKET NUMBER: ORES-5003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-467-7000  
TELEFAX: 202-467-7176  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-649-991-25  
Query Match 48.0%; Score 19.2; DB 4; Length 54;  
Best Local Similarity 75.0%; Pred. No. 13;

Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 9 agtagcgccaagcaaatcaacgtaagaaca 40  
||||| | ||||| ||| ||||| |  
Db 18 AGTAGAGCAAGCAAAACAAAGTAAGAAAA 49

## RESULT 2

US-08-649-991-18  
; Sequence 18, Application US/08649991  
; Patent No. 5919462

; GENERAL INFORMATION:

; APPLICANT: Narva, Remy

; APPLICANT: Roques, Pierre

; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE

; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR

; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF

; NUMBER OF SEQUENCES: 130

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP

; STREET: 1800 M Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20036-5869

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/649,991

; FILING DATE: 17-MAY-1996

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: FR 9505914

; FILING DATE: 18-MAY-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Adler, Reid G.

; REGISTRATION NUMBER: 30,988

; REFERENCE/DOCKET NUMBER: ORES-5003

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-467-7000

; TELEFAX: 202-467-7176

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 54 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; US-08-649-991-18

Query Match 45.5%; Score 18.2; DB 4; Length 54;  
Best Local Similarity 74.2%; Pred. No. 33;

Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 10 gtagacgccaagcaaatcaacgtaagaaca 40  
||||| | ||||| ||| ||||| |  
Db 19 GTAGAGCAAGCAAAACAAAGTAAGAAAA 49

## RESULT 3

US-08-649-991-26

; Sequence 26, Application US/08649991

; Patent No. 5919462

; GENERAL INFORMATION:

; APPLICANT: Narva, Remy

; APPLICANT: Roques, Pierre

; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE

; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR

; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF

; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1

; NUMBER OF SEQUENCES: 130

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP

; STREET: 1800 M Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20036-5869

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/649,991

; FILING DATE: 17-MAY-1996

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: FR 9505914

; FILING DATE: 18-MAY-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Adler, Reid G.

; REGISTRATION NUMBER: 30,988

; REFERENCE/DOCKET NUMBER: ORES-5003

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-467-7000

; TELEFAX: 202-467-7176

; INFORMATION FOR SEQ ID NO: 26:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 54 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; US-08-649-991-26

Query Match 45.5%; Score 18.2; DB 4; Length 54;  
Best Local Similarity 74.2%; Pred. No. 33;

Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 10 gtagacgccaagcaaatcaacgtaagaaca 40  
||||| | ||||| ||| ||||| |  
Db 19 GTAGAGCAAGCAAAACAAAGTAAGAAAA 49

## RESULT 4

US-08-649-991-120

; Sequence 120, Application US/08649991

; Patent No. 5919462

; GENERAL INFORMATION:

; APPLICANT: Narva, Remy

; APPLICANT: Roques, Pierre

; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE

; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR

; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF

; NUMBER OF SEQUENCES: 130

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP

; STREET: 1800 M Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20036-5869

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/08/649,991
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9505914
; FILING DATE: 18-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler, Reid G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: ORES-5003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-467-7000
; TELEFAX: 202-467-7176
; INFORMATION FOR SEQ ID NO: 120:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-649-991-120

Query Match          45.5%; Score 18.2; DB 4; Length 54;
Best Local Similarity 74.2%; Pred. No. 33;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 10 gtacgcgcaagcaaatcaacgyltaagaaca 40
      |||||  |||||  |||  |||||  |
Db 19 gtacagagacgacaaacaaagtaagaaaa 49

RESULT 5
US-08-649-991-10
; Sequence 10, Application US/08649991
; Patent No. 5919462
; GENERAL INFORMATION:
; APPLICANT: Narwa, Remy
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE
; OPERATING SYSTEM: IBM PC compatible
; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR
; APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF
; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP
; STREET: 1800 M Street, N.W.
; City: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5869
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,991
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9505914
; FILING DATE: 18-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler, Reid G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: ORES-5003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-467-7000
; TELEFAX: 202-467-7176
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
```

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-649-991-10

Query Match          43.0%; Score 17.2; DB 4; Length 54;
Best Local Similarity 73.3%; Pred. No. 85;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 11 tagacgcgaagcaaatcaacgyltaagaaca 40
      |||||  |||||  |||  |||||  |
Db 20 tagacgagacgacaaacaaagtaagaaaa 49

RESULT 6
US-08-649-991-11
; Sequence 11, Application US/08649991
; Patent No. 5919462
; GENERAL INFORMATION:
; APPLICANT: Narwa, Remy
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE
; OPERATING SYSTEM: IBM PC compatible
; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR
; APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF
; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP
; STREET: 1800 M Street, N.W.
; City: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5869
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,991
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9505914
; FILING DATE: 18-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler, Reid G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: ORES-5003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-467-7000
; TELEFAX: 202-467-7176
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-649-991-11

Query Match          43.0%; Score 17.2; DB 4; Length 54;
Best Local Similarity 73.3%; Pred. No. 85;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 11 tagacgcgaagcaaatcaacgyltaagaaca 40
      |||||  |||||  |||  |||||  |
Db 20 tagacgagacgacaaacaaagtaagaaaa 49
```

```
RESULT 7
US-08-649-991-12
; Sequence 12, Application US/08649991
; Patent No. 5919462
; GENERAL INFORMATION:
; APPLICANT: Narwa, Remy
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE
; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR
; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF
; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP
; STREET: 1800 M Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5869
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,991
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9505914
; FILING DATE: 18-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler, Reid G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: ORES-5003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-467-7000
; TELEFAX: 202-467-7176
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-649-991-12

Query Match          43.0%; Score 17.2; DB 4; Length 54;
Best Local Similarity 73.3%; Pred. No. 85;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 11 tagagcgcaagcaatcaacggttaagaaca 40
    ||||| ||||| ||| ||||| |||
Db 20 TAGAGGAGAGGCAAAACAAAGTAAGAAAA 49

RESULT 8
US-08-649-991-13
; Sequence 13, Application US/08649991
; Patent No. 5919462
; GENERAL INFORMATION:
; APPLICANT: Narwa, Remy
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE
; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR
; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF
; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP
; STREET: 1800 M Street, N.W.
; CITY: Washington
```

```
STATE: D.C.
COUNTRY: USA
ZIP: 20036-5869
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/649,991
FILING DATE: 17-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9505914
FILING DATE: 18-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Adler, Reid G.
REGISTRATION NUMBER: 30,988
REFERENCE/DOCKET NUMBER: ORES-5003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-467-7000
TELEFAX: 202-467-7176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 54 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-649-991-13

Query Match          43.0%; Score 17.2; DB 4; Length 54;
Best Local Similarity 73.3%; Pred. No. 85;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 11 tagagcgcaagcaatcaacggttaagaaca 40
    ||||| ||||| ||| ||||| |||
Db 20 TAGAGGAGAGGCAAAACAAAGTAAGAAAA 49

RESULT 9
US-08-649-991-14
; Sequence 14, Application US/08649991
; Patent No. 5919462
; GENERAL INFORMATION:
; APPLICANT: Narwa, Remy
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE
; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR
; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF
; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP
; STREET: 1800 M Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5869
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,991
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9505914
; FILING DATE: 18-MAY-1995
; ATTORNEY/AGENT INFORMATION:
```

NAME: Adler, Reid G.  
REGISTRATION NUMBER: 30,988  
REFERENCE/DOCKET NUMBER: ORES-5003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-467-7000  
TELEFAX: 202-467-7176  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-649-991-14

Query Match 43.0%; Score 17.2; DB 4; Length 54;  
Best Local Similarity 73.3%; Pred. No. 85;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 11 tagacgccaagcaaatcaacgtaagaaca 40  
||||| ||||| ||| ||||| ||  
Db 20 TAGAGGAAGACCAAAACAAAGTAGAAGAAA 49

RESULT 10  
US-08-649-991-15  
Sequence 15, Application US/08649991  
Patent No. 5919462  
GENERAL INFORMATION:  
APPLICANT: Narwa, Remy  
TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE  
TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR  
APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF  
TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP  
STREET: 1800 M Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036-5869  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/649,991  
FILING DATE: 17-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9505914  
FILING DATE: 18-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Adler, Reid G.  
REGISTRATION NUMBER: 30,988  
REFERENCE/DOCKET NUMBER: ORES-5003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-467-7000  
TELEFAX: 202-467-7176  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-649-991-15

Query Match 43.0%; Score 17.2; DB 4; Length 54;  
Best Local Similarity 73.3%; Pred. No. 85;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 11 tagacgccaagcaaatcaacgtaagaaca 40  
||||| ||||| ||| ||||| ||  
Db 20 TAGAGGAAGACCAAAACAAAGTAGAAGAAA 49

RESULT 11  
US-08-649-991-17  
Sequence 17, Application US/08649991  
Patent No. 5919462  
GENERAL INFORMATION:  
APPLICANT: Narwa, Remy  
TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE  
TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR  
APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF  
TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP  
STREET: 1800 M Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036-5869  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/649,991  
FILING DATE: 17-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9505914  
FILING DATE: 18-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Adler, Reid G.  
REGISTRATION NUMBER: 30,988  
REFERENCE/DOCKET NUMBER: ORES-5003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-467-7000  
TELEFAX: 202-467-7176  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-649-991-17

Query Match 43.0%; Score 17.2; DB 4; Length 54;  
Best Local Similarity 73.3%; Pred. No. 85;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 11 tagacgccaagcaaatcaacgtaagaaca 40  
||||| ||||| ||| ||||| ||  
Db 20 TAGAGGAAGACCAAAACAAAGTAGAAGAAA 49

RESULT 12  
US-08-649-991-20  
Sequence 20, Application US/08649991  
Patent No. 5919462  
GENERAL INFORMATION:  
APPLICANT: Narwa, Remy  
TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE  
TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR  
APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF  
TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1  
NUMBER OF SEQUENCES: 130  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP  
STREET: 1800 M Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036-5869  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/649,991  
FILING DATE: 17-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9505914  
FILING DATE: 18-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Adler, Reid G.  
REGISTRATION NUMBER: 30,988  
REFERENCE/DOCKET NUMBER: ORES-5003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-467-7000  
TELEFAX: 202-467-7176  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-649-991-17

```

; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE
; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR
; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF
; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP
; STREET: 1800 M Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5869
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,991
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9505914
; FILING DATE: 18-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler, Reid G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: ORES-5003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-467-7000
; TELEFAX: 202-467-7176
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-649-991-20

Query Match          43.0%; Score 17.2; DB 4; Length 54;
Best Local Similarity 73.3%; Pred. No. 85;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 11 tagagcgcaagcaatcaacggttaagaaca 40
    ||||| ||||| ||||| ||||| |||||
Db 20 TAGAGGAGAGCAAAACAAAGTAAGAAAA 49

RESULT 13
US-08-649-991-22
; Sequence 22, Application US/08649991
; Patent No. 5919462
; GENERAL INFORMATION:
; APPLICANT: Narwa, Remy
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE
; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR
; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF
; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP
; STREET: 1800 M Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5869
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,991
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9505914
; FILING DATE: 18-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler, Reid G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: ORES-5003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-467-7000
; TELEFAX: 202-467-7176
; INFORMATION FOR SEQ ID NO: 23:
```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,991
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9505914
; FILING DATE: 18-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler, Reid G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: ORES-5003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-467-7000
; TELEFAX: 202-467-7176
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-649-991-22

Query Match          43.0%; Score 17.2; DB 4; Length 54;
Best Local Similarity 73.3%; Pred. No. 85;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 11 tagagcgcaagcaatcaacggttaagaaca 40
    ||||| ||||| ||||| ||||| |||||
Db 20 TAGAGGAGAGCAAAACAAAGTAAGAAAA 49

RESULT 14
US-08-649-991-23
; Sequence 23, Application US/08649991
; Patent No. 5919462
; GENERAL INFORMATION:
; APPLICANT: Narwa, Remy
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS DERIVED FROM THE
; TITLE OF INVENTION: HIV-1 VIRUS GENOME, CORRESPONDING PEPTIDES AND THEIR
; TITLE OF INVENTION: APPLICATIONS AS REAGENTS FOR EVALUATION OF THE RISK OF
; TITLE OF INVENTION: MATERNOFETAL TRANSMISSION OF HIV-1
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP
; STREET: 1800 M Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5869
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,991
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9505914
; FILING DATE: 18-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler, Reid G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: ORES-5003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-467-7000
; TELEFAX: 202-467-7176
; INFORMATION FOR SEQ ID NO: 23:
```



---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:05:44 ; Search time 1236.38 Seconds

(without alignments)  
-31.472 Million cell updates/sec

Title: US-09-164-714-20

Sequence: 1 gcatgtcgcgtatagatgagctacaagcgtgatttgat 40

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 segs, -486395729 residues

Total number of hits satisfying chosen parameters: 356616

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_com:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl1:\*  
10: gb\_pl2:\*  
11: gb\_pl3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vl:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_or:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vl:\*  
32: gb\_hlg1:\*  
33: gb\_hlg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_p14:\*  
41: gb\_hlg3:\*  
42: gb\_hlg4:\*  
43: gb\_hlg5:\*  
44: gb\_hlg6:\*

45: gb\_hlg7:\*  
46: em\_hlg1:\*  
47: em\_hlg2:\*  
48: em\_hlg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pl5:\*  
52: gb\_hlg6:\*  
53: gb\_hlg9:\*  
54: gb\_hlg10:\*  
55: gb\_hlg11:\*  
56: gb\_hlg12:\*  
57: gb\_hlg13:\*  
58: gb\_hlg14:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.4	43.5	49 5	A00132	A00132 Nucleotide
2	17.4	43.5	54 3	B7BC12YD	X56119 B.Taurus BC
3	16.6	41.5	25 5	A62916	A62916 Sequence 15
4	16	40.0	60 5	I11638	I11638 Sequence 12
5	16	40.0	80 5	A10881	A10881 Synthetic D
6	16	40.0	80 5	A10882	A10882 Synthetic D
7	15.6	39.0	75 9	D50777	D50777 Human mRNA
8	15.2	38.0	35 5	I28730	I28730 Sequence 34
9	15.2	38.0	35 5	I89384	I89384 Sequence 34
10	15.2	38.0	87 34	AGXH487	272092 A.gambiae s
11	14.8	37.0	44 5	A85333	A85333 Sequence 13
12	14.6	36.5	22 5	I59462	I59462 Sequence 11
13	14.6	36.5	33 5	AR014522	AR014522 Sequence
14	14.6	36.5	60 14	SYNANVAA	M60029 Avian neov1
15	14.6	36.5	62 1	JS3PRSEC	M34566 Plasmid pJS
16	14.6	36.5	100 16	RSVRNAL	X67587 Rous sarcom
17	14.4	36.0	45 5	I89301	I89301 Sequence 15
18	14.4	36.0	63 14	SYNDMPRT	M12476 Synthetic P
19	14.4	36.0	70 5	A06245	A06245 Synthetic H
20	14.4	36.0	99 34	DROHIGEO4	D13878 Drosophila
21	14.2	35.5	46 12	MUSIGDJA	L36723 Mus musculu
22	14.2	35.5	57 5	I07481	I07481 Sequence 14
23	14.2	35.5	60 12	MUSIGHTHB	M59872 Mouse Ig sw
24	14.2	35.5	73 34	SUPHISAAC	M25332 Sea urchin
25	14.2	35.5	84 5	AR007052	AR007052 Sequence
26	14.2	35.5	84 5	I74992	I74992 Sequence 19
27	14.2	35.5	84 40	S40179	S40179 pancreatic
28	14.2	35.5	99 5	AR022142	AR022142 Sequence
29	14	35.0	33 5	I14299	I14299 Sequence 7
30	14	35.0	35 5	I86947	I86947 Sequence 36
31	14	35.0	55 5	A87313	A87313 Sequence 38
32	14	35.0	57 5	AR015972	AR015972 Sequence
33	14	35.0	57 5	AR050901	AR050901 Sequence
34	14	35.0	67 12	I33401	I33401 Sequence 1
35	14	35.0	67 12	RNTCRAJ32	Y09184 R.norvegicu
36	14	35.0	72 16	PEOSAB	J02333 Reovirus se
37	14	35.0	77 6	P14TRP	K00354 Bacterioph
38	14	35.0	78 4	CTIURSA	K03165 Duck USA sm
39	14	35.0	79 5	AR062563	AR062563 Sequence
40	14	35.0	79 5	I47287	I47287 Sequence 21
41	14	35.0	89 11	HSVAPOG07	AF030207 Homo sap1
42	13.8	34.5	26 5	AR020758	AR020758 Sequence
43	13.8	34.5	26 5	I34039	I34039 Sequence 13
44	13.8	34.5	39 5	A37820	A37820 Sequence 3
45	13.8	34.5	39 5	AR019386	AR019386 Sequence

#### ALIGNMENTS



```
RESULT 5
A10881/c      A10881      80 bp      DNA      PAT      29-SEP-1993
DEFINITION    Synthetic DNA (MCS) from patent EP0292763.
ACCESSION     A10881
VERSION       A10881.1 GI:490875
KEYWORDS
SOURCE        synthetic construct.
ORGANISM      synthetic construct.
REFERENCE     1 (bases 1 to 80)
AUTHORS      Crause,P.D., Hein,F., Jansen,H.W. and Uhlmann,E.
TITLE        Gene-technological process for the preparation of angiogenines
JOURNAL      Patent: EP 0292763-A 4 30-NOV-1988;
              HOECHST AKTIENGESELLSCHAFT
FEATURES
  source      1..80
              /organism="synthetic construct"
              /db_xref="taxon:32630"
BASE COUNT    17 a      27 c      20 g      16 t
ORIGIN
Query Match   40.0%; Score 16; DB 5; Length 80;
Best Local Similarity 62.5%; Pred. No. 1.5e+04;
Matches 25; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 gcatgtcgcagctagatgacgtacaaagcgctgattggat 40
      ||||| | ||||| | | | ||| ||||| |
Db 79 GCATGCCATGATGATCATGATGCATGCAGAGCTTGGGCT 40

RESULT 6
A10882      A10882      80 bp      DNA      PAT      29-SEP-1993
DEFINITION    Synthetic DNA (MCS) from patent EP0292763.
ACCESSION     A10882
VERSION       A10882.1 GI:489215
KEYWORDS
SOURCE        synthetic construct.
ORGANISM      synthetic construct.
REFERENCE     1 (bases 1 to 80)
AUTHORS      Crause,P.D., Hein,F., Jansen,H.W. and Uhlmann,E.
TITLE        Gene-technological process for the preparation of angiogenines
JOURNAL      Patent: EP 0292763-A 5 30-NOV-1988;
              HOECHST AKTIENGESELLSCHAFT
FEATURES
  source      1..80
              /organism="synthetic construct"
              /db_xref="taxon:32630"
BASE COUNT    16 a      20 c      27 g      17 t
ORIGIN
Query Match   40.0%; Score 16; DB 5; Length 80;
Best Local Similarity 62.5%; Pred. No. 1.5e+04;
Matches 25; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 gcatgtcgcagctagatgacgtacaaagcgctgattggat 40
      ||||| | ||||| | | | ||| ||||| |
Db 2 GCATGCCATGATGATCATGATGCATGCAGAGCTTGGGCT 41

RESULT 7
D50777      D50777      75 bp      mRNA      PRI      10-FEB-1999
LOCUS        Human mRNA for T-cell receptor beta chain V14-D-J2.7-C2 region,
DEFINITION    partial cds.
ACCESSION     D50777
VERSION       D50777.1 GI:1616881
KEYWORDS      T-cell receptor beta chain V14-D-J2.7-C2 region.
SOURCE        Homo sapiens (isolate:YK) male peripheral blood T-lymphocyte cDNA
```

```
ORGANISM      Homo sapiens clone:YK19, YK45.
REFERENCE     Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
              Primates; Catarrhini; Hominiidae; Homo.
AUTHORS      1 (bases 1 to 75)
TITLE        Wang,L.
JOURNAL      Direct Submission
              Submitted (27-JUN-1995) to the DDBJ/EMBL/GenBank databases. L1
              Wang, Japanese Red Cross Central Blood Center, Department of
              Research; 4-1-31 Hiroo, Shibuya-ku, Tokyo 150, Japan
              (Tel:03-5485-6009, Fax:03-3406-7892)
REFERENCE     2 (bases 1 to 75)
AUTHORS      Wang,L., Tadokoro,K., Tokunaga,K., Uchida,S., Mitsuura,S.,
              Bannai,M., Moriyama,S., Takai,K. and Uchi,T.
TITLE        Restricted usage of T-cell receptor Vb gene in post-transfusion
              graft-versus-host disease
JOURNAL      Unpublished (1995)
FEATURES
  source      1..75
              /organism="Homo sapiens"
              /isolate="YK"
              /db_xref="taxon:9606"
              /cell_type="T-lymphocyte"
              /clone="YK19, YK45"
              /sex="male"
              /tissue_type="peripheral blood"
              /gene="TCRB"
              /gene="TCRB"
              /gene="TCRB"
              /note="V14-segment"
              <1..>75
              /gene="TCRB"
              /codon_start=1
              /product="T-cell receptor beta chain V14-D-J2.7-C2 region"
              /protein_id="BA09405.1"
              /db_xref="GI:1616882"
              /translation="ASSSTRGYEQYFGPTRLVTEDL"
              10..22
              /gene="TCRB"
              23..66
              /gene="TCRB"
              /note="J2.7-segment"
              67..>75
              /gene="TCRB"
              /note="C2-segment"
BASE COUNT    18 a      23 c      25 g      9 t
ORIGIN
Query Match   39.0%; Score 15.6; DB 9; Length 75;
Best Local Similarity 63.2%; Pred. No. 2.2e+04;
Matches 24; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 1 gcatgtcgcagctagatgacgtacaaagcgctgattggg 38
      ||| ||||| | | ||||| | | |||
Db 5 GCAGTTCGACAGGAGAACCTACGACAGTACTCGGG 42

RESULT 8
I28730/c     I28730      35 bp      DNA      PAT      06-FEB-1997
LOCUS        Sequence 34 from patent US 5573925.
DEFINITION    I28730
ACCESSION     I28730
VERSION       I28730.1 GI:1819506
KEYWORDS
SOURCE        Unknown.
ORGANISM      Unknown.
REFERENCE     1 (bases 1 to 35)
AUTHORS      Halazonetis,T.D.
TITLE        p53 proteins with altered tetramerization domains
JOURNAL      Patent: US 5573925-A 34 12-NOV-1996;
```



Query Match 36.5%; Score 14.6; DB 5; Length 22;  
Best Local Similarity 81.0%; Pred. No. 5.3e+04;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 18 agctacaagcgctgattgg 38  
|||||  
Db 21 AGCTACAGGCACTCCTGTG 1

RESULT 13  
LOCUS AR014522 33 bp DNA PAT 05-DEC-1998  
DEFINITION Sequence 70 from patent US 5773582.  
ACCESSION AR014522  
VERSION AR014522.1 GI:3971976  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 33)  
AUTHORS Shin,H., Shin,N., Lee,I. and Kang,S.  
TITLE Tumor necrosis factor mutants  
JOURNAL Patent: US 5773582-A 70 30-JUN-1998;  
FEATURES Location/Qualifiers  
source 1..33  
/organism="unknown"

BASE COUNT 6 a 9 c 9 g 9 t  
ORIGIN

Query Match 36.5%; Score 14.6; DB 5; Length 33;  
Best Local Similarity 69.0%; Pred. No. 5.5e+04;  
Matches 20; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 11 gtatgagctacaagcgctgattgg 39  
|||||  
Db 33 GTATGAGCAACAGGCCCTCTATTGCA 5

RESULT 14  
LOCUS SYNANVAA 60 bp DNA SYN 27-APR-1993  
DEFINITION Avian neovirus subgroup A DNA.  
ACCESSION M60029  
VERSION M60029.1 GI:207946  
KEYWORDS  
SOURCE Synthetic viral DNA.  
ORGANISM Artificial construct.  
REFERENCE 1 (bases 1 to 60)  
AUTHORS Olsen,J.C., Bova-Hill,C., Grandgenett,D.P., Quinn,T.P.,  
Manfredi,J.P. and Swanson,R.  
TITLE Rearrangements in unintegrated retroviral DNA are complex and are  
the result of multiple genetic determinants  
JOURNAL J. Virol. 64, 5475-5484 (1990)  
MEDLINE 91012792  
FEATURES Location/Qualifiers  
source 1..60  
/organism="synthetic construct"  
/db\_xref="taxon:32530"

BASE COUNT 16 a 15 c 17 g 12 t  
LTR  
ORIGIN

Query Match 36.5%; Score 14.6; DB 14; Length 60;  
Best Local Similarity 69.0%; Pred. No. 5.8e+04;  
Matches 20; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 9 acgtatgagctacaagcgctgattgg 37  
|||||  
Db 15 ACCTGCATGAAGCAGAGGCTTCATTGG 43

RESULT 15  
LOCUS JS3PRSEC 62 bp DNA BCT 26-APR-1993  
DEFINITION plasmid pJS37 promoter region DNA.  
ACCESSION M34566  
VERSION M34566.1 GI:149149  
KEYWORDS  
SOURCE Plasmid pJS37 DNA, from E.coli (strain C600).  
ORGANISM plasmid pJS37  
REFERENCE 1 (bases 1 to 62)  
AUTHORS Ballester,S., Alonso,J.C., Lopez,P. and Espinosa,M.  
TITLE Comparative expression of the pC194 cat gene in Streptococcus  
JOURNAL pneumoniae, Bacillus subtilis and Escherichia coli  
MEDLINE Gene 86, 71-79 (1990)  
FEATURES 90185250 Location/Qualifiers  
source 1..62  
/organism="plasmid pJS37"  
/plasmid="plasmid pJS37"  
/db\_xref="taxon:2582"  
misc\_signal 21..43  
/note="promoter region PT"  
-35\_signal 21..26  
/note="PT -35 region"  
-10\_signal 43..48  
/note="PT -10 region"  
mRNA 52..>62  
/note="cat mRNA"

BASE COUNT 23 a 3 c 14 g 22 t  
ORIGIN

Query Match 36.5%; Score 14.6; DB 1; Length 62;  
Best Local Similarity 69.0%; Pred. No. 5.8e+04;  
Matches 20; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 12 taatgagctacaagcgctgattgg 40  
|||||  
Db 34 TATATGCTGAATAATGTTAATTGGAT 62

Search completed: June 4, 2000, 16:05:46  
Job time: 27898 sec

\* . . )

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:24:21 ; Search time 322.35 Seconds  
(without alignments)  
31.046 Million cell updates/sec

Title: US-09-164-714-20

Perfect score: 40  
Sequence: 1 gcatgtcgcagctagatgagctacaagcgtattggat 40

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 431286

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.6	46.5	88	1	V39983 Adeno-associated v
2	17.4	43.5	49	1	N70433 Probe b.c.1.2. New
3	17.4	42.5	40	1	V47799 Maize polyporphic
4	17.4	42.5	53	1	V44499 Biotinylated Beque
5	16.8	42.0	98	1	V07390 Synthetic Plasmid
6	16.6	41.5	25	1	T92588 BRCA2 cancer susce
7	16.6	41.5	50	1	X18007 Primer HSA2 for ge
8	16.6	41.5	98	1	V07389 Synthetic Plasmid
9	16.4	41.0	52	1	T73522 Primer RH104 for m
10	16.4	41.0	52	1	T73541 Primer RH104 for m
11	16.4	40.0	60	1	Q49524 Antlerin-2 (OT-2)
12	15.8	39.5	96	1	T49274 HVRI region of E2
13	15.6	39.0	39	1	X05915 HSV VP16 minimal a
14	15.6	39.0	79	1	X05917 HSV VP16 minimal a
15	15.4	38.5	30	1	V12825 Primer for yeast 5
16	15.2	38.0	35	1	T32845 Tumour suppressor
17	15.2	37.5	96	1	T49275 HVRI region of E2
18	14.8	37.0	38	1	N93056 Oligonucleotide fo
19	14.8	37.0	44	1	V57479 Cytochrome P450ox
20	14.8	37.0	71	1	T89124 Human apoptosis in
21	14.8	37.0	88	1	O44249 Probe #4 amplified
22	14.8	37.0	88	1	O44252 VHL disease gene p
23	14.6	36.5	22	1	Q78903 Primer MU-1 for vo
24	14.6	36.5	22	1	T90934 Oligonucleotide SR
25	14.6	36.5	35	1	X15164 PCR primer #1. Scr
26	14.6	36.5	50	1	X02203 Staphylococcus aur
27	14.6	36.5	75	1	V76210 C-C chemokine rece
28	14.4	36.0	25	1	V07412 Scal methylase gen
29	14.4	36.0	45	1	T70165 Bacillus thuringie
30	14.4	36.0	46	1	O68522 crlyc gene primer
31	14.4	36.0	48	1	T09117 PCMV-Fab105-Protam
32	14.4	36.0	53	1	V34508 53 bp intronic seq
33	14.4	36.0	53	1	V34508 Sequence of human
34	14.4	36.0	80	1	Q20226

35	14.4	36.0	80	1	O08622 HPV-16 capture o11
36	14.4	36.0	84	1	X32350 Artificial NS4 mos
37	14.2	35.5	24	1	V27907 Rat GALR3 third tr
38	14.2	35.5	65	1	T39537 Nucleotide used in
39	14.2	35.5	68	1	V52996 Oligonucleotide us
40	14.2	35.5	68	1	V52997 Oligonucleotide us
41	14.2	35.5	99	1	V42880 RNA aptamer bindin
42	14.2	35.5	100	1	T30900 Primer 30 for 95 k
43	14.2	35.0	33	1	O84735 Primer to clone th
44	14.2	35.0	33	1	T03317 Murine IgG heavy c
45	14.2	35.0	35	1	Q36177 Mutagenic primer C

## ALIGNMENTS

RESULT 1	
V39983/C	
ID	V39983 standard; DNA; 88 BP.
AC	V39983;
DT	05-OCT-1998 (first entry)
DE	Adeno-associated virus cap gene pSub201 primer binding region.
KM	Adeno-associated virus; AAV; adenovirus; vector-mediated delivery;
KW	liver; viral particle; tumour; expression; blood disease; haemophilia;
KM	metabolic disease; familial hypercholesterolaemia; primer; ss.
OS	Synthetic.
PN	WO9824479-A1.
PD	11-JUN-1998.
PF	02-DEC-1997; U21398.
PE	25-JUN-1997; US-882044.
PR	02-DEC-1996; US-032506.
PA	(SOMA-) SOMATIX THERAPY CORP.
PA	(UNIW ) UNIV WASHINGTON.
PI	Cohen L, Danos O, Kay M, Snyder R, Thompson AR;
DR	WPI; 98-333055/29.
PT	Expression of polynucleotide(s) in mammals - by administering viral
PT	particles comprising recombinant adeno-associated virus to liver
PT	cells; used for, e.g. treating liver tumours
PS	Example 7; Fig 5; 63pp; English.
CC	Methods have been developed for: (1) expressing a polynucleotide (PN) in
CC	a mammal comprising administering viral particles comprising a
CC	recombinant adeno-associated virus (RAV) vector to liver cells of the
CC	mammal, where the RAV vector comprises the PN, such that the PN, or a
CC	portion, is expressed in the mammal; (2) treating a liver disease or
CC	disorder in a mammal, comprising administering a RAV vector to the
CC	liver cells of the mammal; (3) determining the presence of wild-type AAV
CC	and infectious AAV generated by recombination of a helper AAV and a
CC	vector AAV containing a transgene in a sample of RAVV, where the vector
CC	AAV contains nucleotide sequences or has an order of nt sequences
CC	different from that of wild-type AAV. The RAVV vectors may be used in the
CC	preparation of therapeutics useful for treating a disease or disorder in
CC	a mammal or for gene therapy. The methods permit expression of diffusible
CC	polypeptides in the liver which provides access to the circulation and
CC	permits systemic delivery of therapeutic proteins and polypeptides. They
CC	can be used for the treatment of blood diseases, e.g. haemophilia,
CC	metabolic diseases, e.g. familial hypercholesterolaemia, liver specific
CC	diseases, e.g. glycogen storage diseases or liver tumours. The present
CC	sequence represents an oligonucleotide from the present invention.
SQ	Sequence 88 BP; 22 A; 22 C; 20 G; 24 T;
Query Match	46.5%; Score 18.6; DB 1; Length 88;
Best local similarity	72.7%; Pred. No. 36;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;	
QY	1 gcatgtcgcagctagatgagctacaagcgtatg 33
DB	82 GCATGCTACGTAGATAGTAGCATGGCGGT 50
RESULT 2	
N70433	
ID	N70433 standard; DNA; 49 BP.

AC N70433:  
 DT 16-FEB-1991 (first entry)  
 DE Probe b.c.1.2.  
 KM Probe b.c.1.2.; male-specific probe; ruminant; spermatazoa;  
 OS Bos taurus.  
 PN EP-235046-A.  
 PD 02-SEP-1987.  
 PF 27-FEB-1987; 400434.  
 PR 28-FEB-1986; FR-002811.  
 PR 09-SEP-1986; FR-012616.  
 PA (INRG ) INST NAT RECH AGRON.  
 PA (INSP ) INST PASTEUR.  
 PA (COMS ) COMMISSARIAT A L'ENERGIE ATOMIQUE.  
 PI Bishop C, Colinet C, Fellous M, Kirszenbau M, Vainan M;  
 DR WPI: 87-244362/35.  
 PT New DNA hybridisation probes - specific to ruminant male genome  
 CC Disclosure: page 5; 20pp: french.  
 CC The probe is specific for the male genome of ruminants (esp.  
 CC cattle). It has a hybridisation profile, as determined by  
 CC hybridisation with EcoRI-digested male genomic DNA, with at least  
 CC one band specific to the male genome of the ruminant. The probe is  
 CC repeated at least 2000 times in the bovine male genome. The probe  
 CC is useful for sexing ruminant embryos or foetuses, or for  
 CC monitoring the proportion of spermatozoa carrying the Y chromosome,  
 CC esp. for separating X- or Y-bearing spermatozoa.  
 SQ Sequence 49 Bp; 10 A; 10 C; 16 G; 13 T;

Query Match 43.5%; Score 17.4; DB 1; Length 49;  
 Best Local Similarity 77.8%; Pred. No. 1e+02;  
 Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 8 gacgtagatgagctacaaggcgtgatt 34  
 ||| ||||| ||| ||||| ||| ||  
 DB 12 GACCGAGATGAGCTCTCCACGAGGTGT 38

RESULT 3  
 V47799/c  
 ID V47799 standard; DNA; 40 Bp.  
 AC V47799;  
 DT 14-OCT-1998 (first entry)  
 DE Maize polymorphic site oligonucleotide marker Wx1-G2/G6-5.  
 KM Maize; marker: probe; PCR primer; polymorphism; vegetal sequence;  
 KW polymorphic site; corn; graminæ species; ss.  
 OS Synthetic.  
 OS zea sp.  
 PN WO9830717-A2.  
 PD 16-JUL-1998.  
 PF 02-DEC-1997; E07134.  
 PR 02-DEC-1996; US-032069.  
 PA (BIOC-) BIOCEM SA.  
 PI Murigneux A;  
 DR WPI: 98-399160/34.  
 PT Vegetal sequences including single nucleotide polymorphism - useful,  
 PT e.g. to determine polymorphisms in plants, determine strain in plant  
 PT breeding and to correlate polymorphisms with phenotypic traits  
 PT Claim 2; Page 13; 32pp: English.  
 CC The present invention describes a nucleic acid segment comprising at  
 CC least 10 contiguous nucleotides from a vegetal acid segment including a  
 CC polymorphic site which is a single nucleotide polymorphism (SNP), or the  
 CC complement of the segment. Also described are: (1) an allele-specific  
 CC oligonucleotides hybridising to segment, or their complements, and (2) a  
 CC method of analysing nucleic acids from a subject, by determining if a  
 CC base is occupying any one (or a set) of polymorphic sites in 261  
 CC sequences derived from six maize lines (see V47701 to V47961). The  
 CC segments are useful in fingerprint analysis in plants to determine which  
 CC polymorphisms are present, which strain a plant belongs to and to  
 CC distinguish between strains. The polymorphisms may correlate with  
 CC phenotypic traits (e.g. plant growth rate or crop yield), and the  
 CC segments are useful to determine the presence/absence of specific  
 CC polymorphisms correlating with the existence/absence of particular  
 CC traits. The segments are also useful in marker assisted back-cross

CC techniques to select plants with a higher percentage of recurrent parent  
 CC in a back-cross population. Segments incorporate SNPs which occur more  
 CC frequently than other polymorphism types and are therefore more likely to  
 CC be located close to genetic loci of interest; different forms of  
 CC characterised SNPs are also often easier to detect than other  
 CC polymorphism types.  
 SQ Sequence 40 Bp; 7 A; 17 C; 9 G; 7 T;

Query Match 42.5%; Score 17; DB 1; Length 40;  
 Best Local Similarity 69.7%; Pred. No. 1.4e+02;  
 Matches 23; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 8 gacgtagatgagctacaaggcgtgattgatt 40  
 ||| ||||| ||||| ||||| |||||  
 DB 35 GCCCGAGATGAGCTCTCGCGCTAGTAGGGGCT 3

RESULT 4  
 V44499/c  
 ID V44499 standard; DNA; 53 Bp.  
 AC V44499;  
 DT 20-OCT-1998 (first entry)  
 DE Biotinylated sequencing template.  
 KM DNA sequencing; heterogeneous population; primer; ss.  
 OS Synthetic.  
 PN WO9831831-A1.  
 PD 23-JUL-1998.  
 PR 15-JAN-1998; G00130.  
 PR 15-JAN-1997; GB-000760.  
 PA (BRAX-) BRAX GENOMICS LTD.  
 PI Schmidt G, Thompson AH;  
 DR WPI: 98-427563/36.  
 PT Sequencing of DNA - using an array of hybridisable probes comprising  
 PT a label cleavably attached to a known sequence of predetermined  
 PT length  
 PS Disclosure: Page 25; 51pp: English.  
 CC This sequence is used in a method for sequencing DNA. The method allows  
 CC a heterogeneous population of nucleic acid fragments to be sequenced  
 CC simultaneously. The method does not require traditional gel methods to  
 CC acquire sequence information. Since the entire process takes place in  
 CC solution and is an iterative process, the steps involved could be  
 CC performed by a liquid-handling robot.  
 SQ Sequence 53 Bp; 17 A; 8 C; 12 G; 16 T;

Query Match 42.5%; Score 17; DB 1; Length 53;  
 Best Local Similarity 69.7%; Pred. No. 1.5e+02;  
 Matches 23; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 3 atctcagctagatgagctacaaggcgtgatt 35  
 ||||| ||| ||| ||| ||||| |||||  
 DB 34 ATGTCACCTTACATAACATACAGATCGATT 2

RESULT 5  
 V07390/c  
 ID V07390 standard; DNA; 98 Bp.  
 AC V07390;  
 DT 13-OCT-1998 (first entry)  
 DE Synthetic plasmid construction oligonucleotide.  
 KM Mutant; glucocorticoid receptor; plasmid; constitutively active;  
 KW GR; gene expression; regulation; gene therapy; steroid; treatment;  
 OS Asthma; arthritis; oligonucleotide; ss.  
 OS Synthetic.  
 PN WO9818925-A2.  
 PD 07-MAY-1998.  
 PR 28-OCT-1997; U19607.  
 PR 29-OCT-1996; US-029964.  
 PA (BAYU ) BAYIOR COLLEGE MEDICINE.  
 PA (GENE-) GENEMEDICINE INC.  
 PI Kittle JD, Ledebur HC, Omalley B, Tsai M, Tsai SY,  
 PI Wang Y;



DR WPI: 98-272220/24.  
PT Modified glucocorticoid receptors able to bind non-natural ligands -  
PT for regulating gene expression in gene therapy without requiring  
PT exogenous steroid(s), particularly for treatment of arthritis and  
PT asthma  
PS Example 17: Page 70; 124pp; English.  
CC The sequence is that of an oligonucleotide which was used in the  
CC construction of a plasmid expressing a constitutively active mutant  
CC glucocorticoid receptor (GR) protein.  
SQ Sequence 98 BP; 23 A; 26 C; 26 G; 23 T;

Query Match 42.0%; Score 16.8; DB 1; Length 98;  
Best Local Similarity 66.7%; Pred. No. 2e+02; 12; Indels 0; Gaps 0;  
Matches 24; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 5 gtccagctagatgagctacaagcgctgattggat 40  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 GGCCGGTTATATCTAGAACGGGTGATTCGGAT 36

RESULT 6  
T92588/c  
ID T92588 standard; DNA; 25 BP.  
AC T92588:  
DT 04-FEB-1998 (first entry)  
DE BRCA2 cancer susceptibility gene exon 25B PCR primer F for SSCP.  
KW BRCA2 cancer susceptibility gene; breast cancer; ovarian cancer;  
KW gene therapy; prostate cancer; colorectal cancer; ocular melanoma;  
KW leukemia; human; single stranded conformation polymorphism test;  
KW SSCP; PCR primer; ss.  
OS Synthetic.  
OS Homo sapiens.  
PN GB2307477-A.  
PD 28-MAY-1997.  
PF 25-NOV-1996; 024453.  
PR 28-AUG-1996; GB-017961.  
PR 23-NOV-1995; GB-023959.  
PR 14-DEC-1995; GB-023535.  
PA (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.  
PA (UYDU-) UNIV DUKE.  
PI Ashworth A, Futreal PA, Stratton MR, Wooster RF;  
DR WPI: 97-261854/24.  
PT Nucleic acid molecules comprising part or all of the BRCA2 cancer  
PT susceptibility gene - useful for diagnosis, prognosis or therapeutic  
PT treatment of cancer.  
PS Example 1: Fig 8; 124pp; English.  
CC The present sequence represents a PCR primer for single stranded  
CC conformation polymorphism testing of the BRCA2 cancer susceptibility  
CC gene. The nucleic acid molecule can be used to construct probes for  
CC screening cDNA or genomic libraries, sequencing positive clones  
CC obtained, and assembling the full length BRCA2 sequence. The BRCA2  
CC nucleic acid molecules and proteins are useful in a method of medical  
CC treatment, preferably gene therapy, especially for treating cancer,  
CC where the cancer is female or male breast cancer, ovarian, prostate or  
CC colorectal cancer, ocular melanoma or leukaemia. In particular  
CC antisense oligonucleotides capable of hybridising to the BRCA2 nucleic  
CC acid, pre-mRNA or mature mRNA are used so that the expression of the  
CC BRCA2 nucleic acid is reduced or prevented. The nucleic acid molecules  
CC are also useful in a method for diagnosing susceptibility or  
CC predisposition to cancer in a patient. The nucleic acid molecules are  
CC used to design probes or primers for PCR to determine or detect the  
CC presence of mutations in a sample of nucleic acid from a patient. The  
CC BRCA2 promoter region is useful for screening for substances which  
CC modulate the expression of nucleic acid under control of the promoter.  
CC Antibodies are used to determine the presence, amount or location in a  
CC cell of a BRCA2 polypeptide or its mutant forms. The polypeptides are  
CC used to screen for binding partners, these are useful to screen for  
CC substances which mimic the activity of BRCA2 polypeptide, which can be  
CC used as cancer therapeutics.  
SQ Sequence 25 BP; 7 A; 8 C; 2 G; 8 T;

Query Match 41.5%; Score 16.6; DB 1; Length 25;  
Best Local Similarity 82.6%; Pred. No. 1.9e+02; 4; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 18 agctacaagcgctgattggat 40  
| | | | | | | | | | | | | | | | | | | | | |  
Db 24 AGTAGAAGCGCTGATTTGGCAT 2

RESULT 7  
X18007  
ID X18007 standard; DNA; 50 BP.  
AC X18007:  
DT 21-MAY-1999 (first entry)  
DE Primer HSA2 for generating plasmid pAd/L420-HSA.  
KW Recombinant; adenovirus; vector; interleukin; primer; PCR; amplification;  
KW cancer; cytostatic activity; tumour; ss.  
OS Synthetic.  
OS Mus sp.  
PN W09902685-A1.  
PD 21-JAN-1999.  
PF 13-JUL-1998; NL0406.  
PR 11-JUL-1997; EP-202167.  
PA (INTR-) INTROGENE BV.  
PI Bout A, Draijer-Van Der Kaaden ME;  
DR WPI: 99-120886/10.  
PT Treating tumours with a composition having interleukin-3 activity -  
PT particularly using a recombinant adenoviral vector encoding that  
PT activity  
PS Example 2: Page 26; 79pp; English.  
CC The invention relates to a pharmaceutical composition, comprising a  
CC recombinant adenoviral vector encoding interleukin-3 (IL-3) activity.  
CC The primers X21500 and X18007 were used to construct the recombinant  
CC plasmid pAd/L420-HSA by PCR amplifying the murine HSA gene.  
CC The recombinant adenoviral vector encoding IL-3 activity can be used  
CC in the manufacture of a pharmaceutical composition for the systemic  
CC treatment of cancer. A composition comprising IL-3 activity is used  
CC with a composition comprising cytostatic activity, to treat tumours.  
SQ Sequence 50 BP; 17 A; 8 C; 11 G; 14 T;

Query Match 41.5%; Score 16.6; DB 1; Length 50;  
Best Local Similarity 64.1%; Pred. No. 2.1e+02; 14; Indels 0; Gaps 0;  
Matches 25; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 gcatgctacgtagatgagctacaagcgctgattggat 39  
| | | | | | | | | | | | | | | | | | | | | |  
Db 12 GCTGTGCACTGCATCTACTACAGTAGAGATGTAGAA 50

RESULT 8  
V07389/c  
ID V07389 standard; DNA; 98 BP.  
AC V07389:  
DT 13-OCT-1998 (first entry)  
DE Synthetic plasmid construction oligonucleotide.  
KW Mutant; glucocorticoid receptor; plasmid; constitutively active;  
KW GR; gene expression; regulation; gene therapy; steroid; treatment;  
KW asthma; arthritis; oligonucleotide; ss.  
OS Synthetic.  
OS W09818925-A2.  
PN W09818925-A2.  
PD 07-MAY-1998.  
PF 28-OCT-1997; U19607.  
PR 29-OCT-1996; US-029964.  
PA (BAYU ) BAYLOR COLLEGE MEDICINE.  
PA (GENE-) GENEMEDICINE INC.  
PI Kittle JD, Ledebur HC, Omalley B, Tsai M, Tsai SY,  
PI Wang Y;  
DR WPI: 98-272220/24.  
PT Modified glucocorticoid receptors able to bind non-natural ligands -  
PT for regulating gene expression in gene therapy without requiring  
PT exogenous steroid(s), particularly for treatment of arthritis and  
PT asthma









**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 13:53:37 ; Search time 4521.53 Seconds  
(without alignments)  
35.857 Million cell updates/sec

Title: US-09-164-714-20

Perfect score: 40

Sequence: 1 gcatgtcgaactagatgagctacaagcgatggat 40

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 segs, 202661650 residues

Total number of hits satisfying chosen parameters: 119968

Minimum DB seq length: 0

Maximum DB seq length: 100

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: gb\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_est48:\*  
83: gb\_est49:\*  
84: gb\_est50:\*  
85: gb\_est51:\*  
86: em\_est52:\*  
87: em\_est53:\*  
88: em\_est54:\*  
89: gb\_est55:\*  
90: gb\_est56:\*  
91: gb\_est57:\*  
92: gb\_est58:\*  
93: gb\_est59:\*  
94: gb\_est60:\*  
95: em\_est61:\*  
96: em\_est62:\*  
97: em\_est63:\*  
98: em\_est64:\*  
99: em\_est65:\*  
100: em\_est66:\*  
101: em\_est67:\*  
102: gb\_est68:\*  
103: gb\_est69:\*  
104: em\_est70:\*  
105: gb\_est71:\*  
106: gb\_est72:\*  
107: gb\_est73:\*  
108: gb\_est74:\*  
109: gb\_est75:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result

Query

SUMMARIES

No.	Score	Match	Length	DB	ID	Description
1	17.2	43.0	68	50	F26502	F26502 HSPD14000 H
2	17	42.5	94	40	AA967015	AA967015 ua39bl1.r
3	16.4	41.0	77	37	AA730273	AA730273 nz94b05.s
4	16.2	40.5	86	33	AA431534	AA431534 zw7ch06.r
5	16.2	40.5	94	36	AA641353	AA641353 nr91e01.s
6	16	40.0	66	40	AA912094	AA912094 o157g09.s
7	16	40.0	82	24	H66370	H66370 yu14b03.r1
8	16	40.0	82	30	AA212741	AA212741 mw83c10.r
9	15.8	39.5	58	23	HA9949	HA9949 yo25c03.r1
10	15.8	39.5	85	59	AT767746	AT767746 wh3c07.x
11	15.8	39.5	96	23	R84750	R84750 yq28a09.r1
12	15.8	39.5	100	26	w77250	w77250 me65e02.r1
13	15.6	39.0	71	45	AT360727	AT360727 qx98d03.x
14	15.4	38.5	67	42	AT159257	AT159257 vz88b08.r
15	15.4	38.5	70	74	AA215038	AA215038 up03h04.y
16	15.4	38.5	98	50	AT689524	AT689524 tx13d01.x
17	15.2	38.0	78	44	AT267735	AT267735 ap64a02.x
18	15.2	38.0	94	27	AA000065	AA000065 mg27g09.r
19	15.2	38.0	95	49	AT664297	AT664297 ue84a12.r
20	15.2	38.0	99	62	AT903060	AT903060 OY-BT022-774264.yc56g05.s1
21	15	37.5	52	21	T74264	T74264 yc56g05.s1
22	15	37.5	52	37	AA723687	AA723687 ah85c05.s
23	15	37.5	82	44	AT324951	AT324951 ml74h09.x
24	15	37.5	84	42	AF027883	AF027883 AF027883
25	15	37.5	88	21	T99680	T99680 ye67d08.r1
26	15	37.5	94	38	AA765118	AA765118 n261c10.s
27	15	37.5	97	49	AT656392	AT656392 tt50e11.x
28	15	37.5	97	62	AT928344	AT928344 wo44g01.x
29	15	37.5	100	81	AA453970	AA453970 zeh10745
30	14.6	36.5	55	30	AA214324	AA214324 zq91e11.s
31	14.6	36.5	67	40	AA936694	AA936694 o179d04.s
32	14.6	36.5	73	24	D18181	D18181 MUGS00446
33	14.6	36.5	77	45	AT393278	AT393278 tg14e07.x
34	14.6	36.5	80	20	T11114	T11114 hbc700.Huma
35	14.6	36.5	89	30	AA266257	AA266257 mz67a02.r
36	14.6	36.5	91	27	AA039440	AA039440 zk39b09.r
37	14.6	36.5	92	42	AT1121560	AT1121560 uc43c09.r
38	14.6	36.5	95	25	U44258	U44258 ENM44258.As
39	14.6	36.5	100	64	AA063099	AA063099 SWMAC443
40	14.4	36.0	55	39	AA862559	AA862559 oh44c03.s
41	14.4	36.0	58	63	AT940991	AT940991 sb82h08.y
42	14.4	36.0	70	26	W47924	W47924 mc84e09.r1
43	14.4	36.0	82	39	AA845785	AA845785 ak77h10.s
44	14.4	36.0	87	73	AM168701	AM168701 x194e06.x
45	14.4	36.0	88	37	AA699901	AA699901 zi61e02.s

## ALIGNMENTS

RESULT 1  
LOCUS F26502 68 bp mRNA EST 13-MAY-1998  
DEFINITION HSPD14000 HM3 Homo sapiens cDNA clone s4000056B08, mRNA sequence.  
ACCESSION F26502  
VERSION F26502.1 GI:4812128  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Pandolfo,D., Toppo,S., Trevisan,S., Scarso,S. and Valle,G.  
TITLE Identification of 4370 expressed sequence tags from a 3'-end-specific cDNA library of human skeletal muscle by DNA sequencing and filter hybridization

JOURNAL Genome Res. 6 (1), 35-42 (1996)  
MEDLINE 96276048  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948143.  
Contact: Valle G.

CRIBI Biotechnology Centre  
University of Padua  
Via Trieste 75, 35121 Padua, Italy  
ABI Chromatograms and other information are available on WWW at  
<http://grupp.bio.unipd.it>.

FEATURES  
source Location/Qualifiers

1. 68  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="s4000056B08"  
/clone\_lib="HM3"  
/sex="female"

/tissue\_type="pectoral muscle (after mastectomy)"  
/note="Vector: pCDNMT1 (Invitrogen); Site\_1: BstXI;  
site\_2: NotI; The library was constructed by G.  
lanfranchi. This library is not subtracted nor normalized.

The first strand cDNA was primed with a biotinylated  
oligo-dT-NotI primer

(5'-Diotin-AACCGGCTCGAGCGCCCTTTTCTTTTCTTTTCTTTT-3'). The  
ds cDNA was sonicated and size-selected in the range  
350-550 bp. The 3' specific fragments were selected by  
streptavidin coated magnetic beads, ligated to  
non palindromic BstXI adapters, NotI digested and  
directionally cloned into BstXI-NotI cut pCDNMT1 vector."

BASE COUNT 22 a 19 c 18 g 9 t

## ORIGIN

Query Match 43.0%; Score 17.2; DB 50; Length 68;  
Best Local Similarity 73.3%; Pred. No. 3e+03; 8; Indels 0;  
Matches 22; Conservative 0; Mismatches 8; Gaps 0;

OY 2 catgtgcacgtatgacgtacacggcggtg 31  
DB 2 CAAGGCCATGACAGATGTACAGCGCGT 31

RESULT 2  
LOCUS AA967015/c 94 bp mRNA EST 19-MAY-1998  
DEFINITION ua39bl1.r1 Soares\_mammary\_gland\_NBMG Mus musculus cDNA clone  
IMAGE:1349085 5' similar to SW:THIO.CYAC P37395 THIOREDOXIN.;  
KEYWORDS mRNA sequence.  
ACCESSION AA967015  
VERSION AA967015.1 GI:3140908  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geislenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.

TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Jan 14, 1998 this sequence version replaced gi:1796941.  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: [moussest@watson.wustl.edu](mailto:moussest@watson.wustl.edu)  
This clone is available royalty-free through LUNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:697877

Trace considered overall poor quality  
Possible reversed clone: similarity on wrong strand  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 1.



FEATURES		SOURCE		Location/Qualifiers	
		1. .94		/organism="Mus musculus"	
		/strain="C57BL/6J"		/db_xref="taxon:10090"	
		/clone="IMAGE:1349085"		/clone_lib="Soares_mammary_gland_NbMNG"	
		/sex="male"		/tissue_type="mammary_gland"	
		/dev_stage="4 weeks"		/lab_host="DH10B"	
		/note="Organ: mammary gland: Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5',		TGTTCACAACTGTGAATGTGGAGCGCGCCGCAATGCTTTTTTTTTTTTTTTTTTTT	
		T 3']:: double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."			
BASE COUNT		26 a	22 c	23 g	23 t
ORIGIN					
Query Match	Best Local Similarity	42.5%;	Score 17;	DB 40;	Length 94;
Matches	20; Conservative	80.0%;	Pred. No. 3.9e+03;	Mismatches	5; Indels
		0;		Gaps	0;
OY	13 agatagctacaagcgctgattgg 37				
	1 1111 111111 1111 1111				
Db	30 AAATGACCAACAGAGAGTGG 6				
RESULT	3				
AA730273	LOCUS	AA730273	77 bp	mRNA	EST
	DEFINITION	n294b05.s1 NCI_CGAP_GCB1 Homo sapiens CDNA clone IMAGE:1303089 3',			
	ACCESSION	AA730273			
	VERSION	AA730273.1	GI:2753485		
	KEYWORDS	EST.			
	SOURCE	human.			
	ORGANISM	Homo sapiens			
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
		Eutheria; Primates; Catarrhini; Homnidae; Homo.			
	REFERENCE	1 (bases 1 to 77)			
	AUTHORS	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .			
	TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index			
	JOURNAL	Unpublished (1997)			
	COMMENT	On Sep 12, 1996 this sequence version replaced gi:1400835.			
		Contact: Robert Strausberg, Ph.D.			
		Tel: (301) 496-1550			
		Email: Robert.Strausberg@nih.gov			
		Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,			
		Ph.D., Gerald Marti, M.D.			
		CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima			
		Bonaldo, Ph.D.			
		CDNA Library Arrayed by: Greg Lennon, Ph.D.			
		DNA Sequencing by: Washington University Genome Sequencing Center			
		Clone distribution: NCI-CGAP clone distribution information can be			
		found through the I.M.A.G.E. Consortium/LNL at:			
		<a href="http://www-bio.lnl.gov/bbrp/image/image.html">www-bio.lnl.gov/bbrp/image/image.html</a>			
FEATURES		Insert Length: 805 Std Error: 0.00			
		Seq primer: ~40m13 fwd. ET from Amersham			
		High quality sequence stop: 68.			
SOURCE		Location/Qualifiers			
		1. .77			
		/organism="Homo sapiens"			
		/db_xref="taxon:9606"			

FEATURES	Source
BASE COUNT	25 a 6 c 17 g 29 t
ORIGIN	<p>           /clone="IMAGE:1303089"            /clone_lib="NCI-CGAP_GCB1"            /tissue_type="germinal center B cell"            /lab_host="DH10B"            /note="Vector: pRT73D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer            [5'-GTGACCAATCGAAGGAGGAGCGCCGCTCATTTTTTTTTTTTTTT-3'] Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."         </p>
Query Match	41.0%; Score 16.4; DB 37; Length 77;
Best Local Similarity	76.9%; Pred. No. 6.4e+03;
Matches	20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db	15 atgagctacaagcgctgattggagat 40                           30 ATGACTTAAGGACTGATTGGCTT 55
RESULT	4
LOCUS	AA431534 86 bp mRNA EST 22-MAY-1997
DEFINITION	zw76h06.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:782171
ACCESSION	AA431534
VERSION	AA431534.1 GI:2115242
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 86) Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
TITLE	WashU-Merck EST Project 1997
JOURNAL	Unpublished (1997)
COMMENT	On May 8, 1995 this sequence version replaced gi:800243. Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -28m13 rev2 ER from Amersham High quality sequence stop: 6. Location/Qualifiers 1. 86 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:782171" /clone_lib="Soares_testis_NHT" /sex="male" /lab_host="DH10B" /note="Vector: pRT73D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5'

TGTTACCATCTGAAGTGGAGCGCCGCCCAATTTTCTTTTCTTTT 3'.

Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to cdt5, and was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 28 a 28 c 17 g 13 t

Query Match 40.5%; Score 16.2; DB 33; Length 86;  
Best Local Similarity 64.9%; Pred. No. 7.9e+03;  
Matches 24; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 1 gcatgcagctagatgactacaagcgctgattgg 37  
I | | | | | I | | | | | I | | | | | I | | | | |  
Db 78 GAATGCACCTCTGCGCTGGATGGCTGTGTTTGG 42

RESULT 5  
AA641353 94 bp mRNA EST 27-OCT-1997  
LOCUS nr191e01.s1 NCI-CGAP Pr25 Homo sapiens cDNA clone IMAGE:1175352 3'

DEFINITION similar to SW:RSS\_HUMAN P46782 40S RIBOSOMAL PROTEIN S5. ;, mRNA sequence.

ACCESSION AA641353  
VERSION AA641353.1 GI:2566603

KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 94)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index

JOURNAL Unpublished (1997)  
COMMENT On Apr 14, 1993 this sequence version replaced gi:693048.

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Suzanne L. Topalian, M.D., Robert K. Bright,  
Ph.D.

CDNA Library Preparation: Stratagene, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/btrp/image/image.html

Trace considered overall poor quality  
Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 1.  
Location/Qualifiers

1..94  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1175352"  
/clone.lib="NCI-CGAP.Pr25"  
/tissue\_type="epithelium (cell line)"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: prostate; Vector: Bluescript SK-; Site:1:  
EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. Normal prostate epithelial cell line (HPV  
immortalized). 5' adaptor sequence: 5' GAATTCGCGACGAG 3'  
3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'  
Average insert size: 1.1 kb."

BASE COUNT 20 a 23 c 26 g 25 t

ORIGIN

Query Match 40.5%; Score 16.2; DB 36; Length 94;  
Best Local Similarity 64.9%; Pred. No. 8.1e+03;

Matches 24; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 4 tctgcagctagatgactacaagcgctgattggat 40  
I | | | | | I | | | | | I | | | | | I | | | | |  
Db 34 TGAGCAGGTGTATGATCTCGAAGCATGCTTACGAT 70

RESULT 6  
AA912094 66 bp mRNA EST 21-DEC-1998  
LOCUS o157g09.s1 NCI-CGAP HN4 Homo sapiens cDNA clone IMAGE:1486816 3'

DEFINITION similar to gb:K00558 TUBULIN ALPHA-1 CHAIN (HUMAN);, mRNA sequence.

ACCESSION AA912094  
VERSION AA912094.1 GI:3051486

KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 66)  
NCI/NIDR-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS National Cancer Institute / National Institute of Dental Research,  
TITLE Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)  
COMMENT On Jan 19, 1998 this sequence version replaced gi:2287030.

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov  
Tissue Procurement: John Ensley, M.D.

CDNA Library Preparation: Stratagene, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/btrp/image/image.html

Trace considered overall poor quality  
Insert Length: 1728 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 1.  
Location/Qualifiers

1..66  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1486816"  
/clone.lib="NCI-CGAP.HN4"  
/tissue\_type="squamous cell carcinoma"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: pharynx; Vector: Bluescript SK-; Site:1:  
EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. Average insert size 1.5 kb. 5' adaptor sequence:  
5' GAATTCGCGACGAG 3' 3' adaptor sequence: 5'  
(GA)10ACTGATCGAGTCTTTTCTTTTCTTTTCTTTT 3'."

BASE COUNT 17 a 16 c 17 g 16 t

ORIGIN

Query Match 40.0%; Score 16; DB 40; Length 66;  
Best Local Similarity 68.8%; Pred. No. 8.9e+03;  
Matches 22; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 6 tcgacgtagatgactacaagcgctgattgg 37  
I | | | | | I | | | | | I | | | | | I | | | | |  
Db 47 TTGACACGTATGATATACGAGCGTCCCTTGG 16

RESULT 7  
H66370 82 bp mRNA EST 18-OCT-1995  
LOCUS yu14b03.r1 Soares fetal liver spleen INFIS Homo sapiens cDNA clone  
DEFINITION IMAGE:233741 5' similar to SP:T26A5.9 CE00788 ;, mRNA sequence.

ACCESSION H66370  
VERSION H66370.1 GI:1025110

	KEYWORDS	EST.
	SOURCE	human.
	ORGANISM	Homo sapiens
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
		Eutheria; Primates; Catarrhini; Homidae; Homo.
	REFERENCE	1 (bases 1 to 82)
	AUTHORS	Hallier,L., Clark,N., Dubnue,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Woldmann,P. and Wilson,R.
TITLE	The Mashu-Merck EST project	
JOURNAL	Unpublished (1995)	
COMMENT	On May 9, 1995 this sequence version replaced gi:803014. Contact: Wilson RK Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Insert size: 1043 High quality sequence starts: 1 High quality sequence stops: 1 Source: IMAGE Consortium, LNLN This clone is available royalty-free through LNLN ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Trace considered overall poor quality Possible reversed clone: similarity on wrong strand Insert Length: 1043 Std Error: 0.00 Seq primer: M13RP1 High quality sequence stop: 1. Location/Qualifiers 1..82 /organism="Homo sapiens" /db_xref="GDB:3768918" /db_xref="taxon:9606" /clone="IMAGE:233741" /clone_lib="Soares fetal liver spleen INFLS" /sex="male" /dev_stage="20 week-post conception fetus" /lab_host="DH10B (ampicillin resistant)" /note-"Organ: Liver and Spleen; Vector: pTR73D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACGTGAAGATTAAATAAAGATCTTTTCTTTTTTTT 3'] , double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pTR73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."	
FEATURES		
SOURCE		
BASE COUNT	16 a 19 c 17 g 26 t	4 others
ORIGIN		
Query Match	40.0%; Score 16; DB 24; Length 82;	
Best Local Similarity	62.9%; Pred. NO. 9.4e+03;	
Matches	22; Conservative	0; Mismatches 13; Indels 0; Gaps 0;
OY	2 catgtcgacgtagatgaagctacaaggcgtatttg 36	
Dd	11 11 11111111 1 11 11 11	
	11 11 11111111 1 11 11 11	
	CAGNTAGAAGTAGATGAAGTGTTGGNGTCATGTG 15	
RESULT		
LOCUS	AA12741/c	
DEFINITION	AA12741 82 bp mRNA EST 18-FEB-1997	
ACCESSION	AA12741	
VERSION	AA12741.1 GI:1811370	
KEYWORDS	EST.	
SOURCE	mouse mouse.	

ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 82) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Getzel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellensberg,K., Stepien,M., Tan,F., Underwood,K., Moore,B., Theissen,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE	The Mashu-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On May 5, 1995 this sequence version replaced gi:797664. Contact: Marra M/Mouse EST Project Mashu-HHMI Mouse EST Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MG1:417002
FEATURES	Trace considered overall poor quality Possible reversed clone: similarity on wrong strand Seq primer: -26ml3 rev2 EF from Amersham High quality sequence stop: 1.
SOURCE	location/Qualifiers 1..82 /organism="Mus musculus" /db_xref="taxon:10090" /clone_1id="IMAGE:677298" /clone_1lb="Soares mouse NML" /rissue_type="Liver" /lab_host="DH10B" /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAAGTGGAAGCGCCGCACAATCTTTTTTTTTTTT 3']; TGTTCACCAATCTGAAAGTGGAAGCGCCGCACAATCTTTTTTTTTTTT 3'; (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed and normalized by Bento Soares and M.Palima BonaIdo."
BASE COUNT	18 a 19 c 22 g 23 t
ORIGIN	
Query Match	40.0%, Score 16; DB 30; Length 82;
Best Local Similarity	79.2%; Pred. No. 9.4e+03;
Matches	19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Oy	13 agatgaactacaagcgctgattg 36                               24 AATGAMCCACAGAAGACTTG 1
RESULT	9
LOCUS	H49949
DEFINITION	yo25go3.r1 Soares adult brain NZB5HB55Y Homo sapiens cDNA clone IMAGE:178996 5' similar to SP:SVL_DISOM P24505 SYNAPTORAGMIN A ; , mRNA sequence.
ACCESSION	H49949
VERSION	H49949.1
KEYWORDS	GI:989790
SOURCE	EST.
ORGANISM	human.
Homio sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
1 (bases 1 to 58)	
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,	







FEATURES  
source

Location/Qualifiers  
1..70  
/organism="Mus musculus"  
/strain="CZECH IT"  
/db\_xref="taxon:10090"  
/clone="IMAGE:2650999"  
/clone\_lib="NCI CGAP Lu30"  
/issue\_type="tumor, metastatic to mammary"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: PCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; transgenic model WNT-1, expression driven by  
MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo  
dT. Library constructed by Life Technologies.  
Investigator providing samples: Gilbert Smith, NIH"  
BASE COUNT 7 a 15 c 32 g 16 t  
ORIGIN

Query Match 38.5%; Score 15.4; DB 74; Length 70;  
Best Local Similarity 76.0%; Pred. No. 1.6e+04;  
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 15 atgagctacaagcggtgatttgga 39  
||| ||| ||| ||| ||| ||| ||| |||  
Db 46 ATGGGCTTCAAGCGAGATCAGGGA 22

Search completed: June 4, 2000, 13:53:40  
Job time: 20964 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2000, 16:09:48 ; Search time 244.64 Seconds  
(without alignments)  
21,253 Million cell updates/sec

Title: US-09-164-714-20

Perfect score: 40  
Sequence: 1 gcatgtcagctagatgaagcacaagcgctattggat 40

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 375860

Minimum DB seq length: 0  
Maximum DB seq length: 100

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

1: /cgn2\_6/prodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/5C.COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/5D.COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/6.COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/PCRUS.COMB.seq:\*  
7: /cgn2\_6/prodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.6	41.5	25	5	US-08-755-587-157 Sequence 157, App
2	16	40.0	60	1	US-08-180-572-12 Sequence 12, App
3	15.2	38.0	35	1	US-08-347-792-34 Sequence 34, App
4	15.2	38.0	35	2	US-08-431-357-34 Sequence 34, App
5	15.2	38.0	35	6	PCT-US95-15353-34 Patent No. 5194596
6	14.8	37.0	88	7	5194596-1 Patent No. 5194596
7	14.8	37.0	88	7	5194596-4 Patent No. 5194596
8	14.8	37.0	88	7	5219739-1 Patent No. 5219739
9	14.8	37.0	88	7	5219739-4 Patent No. 5219739
10	14.6	36.5	22	1	US-08-061-889-11 Sequence 11, App
11	14.6	36.5	22	1	US-08-462-611-11 Sequence 11, App
12	14.6	36.5	22	6	PCT-US94-05378-11 Sequence 11, App
13	14.6	36.5	33	2	US-08-538-875-70 Sequence 70, App
14	14.6	36.5	33	4	US-08-600-783-8 Sequence 8, App
15	14.6	36.5	33	4	US-08-600-783-9 Sequence 9, App
16	14.6	36.5	50	3	US-08-867-579-1 Sequence 1, App
17	14.6	36.5	78	3	US-08-659-251-16 Sequence 16, App
18	14.6	36.5	78	6	PCT-US96-11445-16 Sequence 16, App
19	14.4	36.0	45	2	US-08-569-806-15 Sequence 15, App
20	14.4	36.0	46	2	US-08-433-783-24 Sequence 24, App
21	14.4	36.0	46	3	US-08-337-358-24 Sequence 24, App
22	14.4	36.0	46	6	PCT-US93-12144-24 Sequence 24, App
23	14.4	36.0	46	6	PCT-US95-07537A-24 Sequence 24, App
24	14.4	36.0	46	6	PCT-US95-07537B-24 Sequence 24, App
25	14.4	36.0	48	6	PCT-US95-02140-5 Sequence 5, App
26	14.2	35.5	65	6	PCT-US95-13376-16 Sequence 16, App
27	14.2	35.5	84	1	US-08-433-126A-191 Sequence 191, App

28	14.2	35.5	84	2	US-08-433-124A-191 Sequence 191, App
29	14.2	35.5	84	6	PCT-US96-06059-191 Sequence 191, App
30	14.2	35.5	99	2	US-08-662-335A-10 Sequence 10, App
31	14	35.0	33	1	US-07-862-495-7 Sequence 7, App
32	14	35.0	35	2	US-08-170-290A-36 Sequence 36, App
33	14	35.0	57	1	US-08-096-027-1 Sequence 1, App
34	14	35.0	57	2	US-08-461-725-1 Sequence 1, App
35	14	35.0	57	3	US-08-460-981-1 Sequence 1, App
36	14	35.0	79	1	US-08-384-708A-217 Sequence 217, App
37	14	35.0	79	3	US-08-470-939-14 Sequence 14, App
38	14	35.0	79	6	PCT-US96-09452-14 Patent No. 5194596
39	14	35.0	88	7	5194596-6 Patent No. 5194596
40	13.8	34.5	26	1	US-08-117-362-13 Sequence 13, App
41	13.8	34.5	26	2	US-08-466-924-13 Sequence 13, App
42	13.8	34.5	26	2	US-08-468-347-4 Sequence 4, App
43	13.8	34.5	39	2	US-08-226-264-2 Sequence 2, App
44	13.8	34.5	39	2	US-08-226-264-2 Sequence 2, App
45	13.8	34.5	39	2	US-08-226-264-11 Sequence 11, App

#### ALIGNMENTS

RESULT 1  
US-08-755-587-157/C  
Sequence 157, Application US/08755587  
Patent No. 6045997  
GENERAL INFORMATION:  
APPLICANT: Futreal, Phillip A  
APPLICANT: Wooster, Richard F  
APPLICANT: Ashworth, Alan  
APPLICANT: Stratton, Michael R  
TITLE OF INVENTION: Materials and methods relating to the  
TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer  
NUMBER OF SEQUENCES: 222  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Bell Seltzer Park & Gibson  
STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107  
CITY: Raleigh  
STATE: NC  
COUNTRY: USA  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (ERO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/755,587  
FILING DATE: 25-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9523959.6  
FILING DATE: 23-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9525555.0  
FILING DATE: 14-DEC-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9617961.9  
FILING DATE: 28-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenneth D Sibley  
REGISTRATION NUMBER: 31,665  
REFERENCE/DOCKET NUMBER: 5405-135  
INFORMATION FOR SEQ ID NO: 157:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-755-587-157  
Query Match 41.5%; Score 16.6; DB 5; Length 25;



ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/431,357  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/347,792  
FILING DATE: 28-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: WST5805A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9206  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-431-357-34

Query Match 38.0%; Score 15.2; DB 2; Length 35;  
Best Local Similarity 71.4%; Pred. No. 3.4e+02;

Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 6 tcgacgtagatgaactacaagcgtgat 33  
| | | | | | | | | | | | | | | | | |  
Db 28 TGGGATCGATGAGCTTCCAGGCTCAT 1

RESULT 5  
PCT-US95-15353-34/C  
Sequence 34, Application PC/TUS9515353  
GENERAL INFORMATION:  
APPLICANT: The Wistar Institute of Anatomy  
and Biology  
APPLICANT: Halazonetis, Thanos D.  
TITLE OF INVENTION: p53 Proteins With Altered  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr., PO Box 457  
City: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/15353  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/347,792  
FILING DATE: 28-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/431,357  
FILING DATE: 28-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/456,623

FILING DATE: 01-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: WST580PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9206  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
PCT-US95-15353-34

Query Match 38.0%; Score 15.2; DB 6; Length 35;  
Best Local Similarity 71.4%; Pred. No. 3.4e+02;

Matches 20; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 6 tcgacgtagatgaactacaagcgtgat 33  
| | | | | | | | | | | | | | | | | |  
Db 28 TGGGATCGATGAGCTTCCAGGCTCAT 1

RESULT 6  
5194596-1/C  
Patent No. 5194596  
APPLICANT: TISCHER, EDMUND G.; ABRAHAM, JUDITH A.; FIDES, JOHN  
C.; MITCHELL, RICHARD L.  
TITLE OF INVENTION: PRODUCTION OF VASCULAR ENDOTHELIAL CELL  
GROWTH FACTOR  
NUMBER OF SEQUENCES: 32  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/450,883  
FILING DATE: 14-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 387,545  
FILING DATE: 27-JUL-1989  
SEQ ID NO: 1:  
LENGTH: 88  
5194596-1

Query Match 37.0%; Score 14.8; DB 7; Length 88;  
Best Local Similarity 73.1%; Pred. No. 6.2e+02;  
Matches 19; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 gcatgtcagctagatgaactacaag 26  
| | | | | | | | | | | | | | | | | |  
Db 29 GCTGTAGAGCTCATGAACTTCAAG 4

RESULT 7  
5194596-4/C  
Patent No. 5194596  
APPLICANT: TISCHER, EDMUND G.; ABRAHAM, JUDITH A.; FIDES, JOHN  
C.; MITCHELL, RICHARD L.  
TITLE OF INVENTION: PRODUCTION OF VASCULAR ENDOTHELIAL CELL  
GROWTH FACTOR  
NUMBER OF SEQUENCES: 32  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/450,883  
FILING DATE: 14-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 387,545  
FILING DATE: 27-JUL-1989  
SEQ ID NO: 4:  
LENGTH: 88  
5194596-4

Query Match 37.0%; Score 14.8; DB 7; Length 88;  
Best Local Similarity 73.1%; Pred. No. 6.2e+02;  
Matches 19; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 gcatgtcagcagtagatgagctacaag 26  
||| ||||| ||||| ||| |||||  
DB 29 GCTGTAGACGTCCATGAACTTCAAG 4

RESULT 8  
5219739-1/c

; Patent No. 5219739  
; APPLICANT: TISCHER, EDMUND G.; ABRAHAM, JUDITH A.; FIDES,  
; JOHN C.; MITCHELL, RICHARD L.  
; TITLE OF INVENTION: DNA SEQUENCES ENCODING BVEGF120 AND  
; BVGEF 121 AND METHODS FOR THE PRODUCTION OF BOVINE AND HUMAN  
; VASCULAR ENDOTHELIAL CELL GROWTH FACTORS, BVGEF120 AND BVGEF121  
; NUMBER OF SEQUENCES: 40  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/559,041  
; FILING DATE: 27-JUL-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 450,883  
; FILING DATE: 14-DEC-1989  
; APPLICATION NUMBER: 387,545  
; FILING DATE: 27-JUL-1989  
; SEQ ID NO:1:  
; LENGTH: 88  
5219739-1

Query Match 37.0%; Score 14.8; DB 7; Length 88;  
Best Local Similarity 73.1%; Pred. No. 6.2e+02;  
Matches 19; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 gcatgtcagcagtagatgagctacaag 26  
||| ||||| ||||| ||| |||||  
DB 29 GCTGTAGACGTCCATGAACTTCAAG 4

RESULT 9  
5219739-4/c  
; Patent No. 5219739  
; APPLICANT: TISCHER, EDMUND G.; ABRAHAM, JUDITH A.; FIDES,  
; JOHN C.; MITCHELL, RICHARD L.  
; TITLE OF INVENTION: DNA SEQUENCES ENCODING BVGEF120 AND  
; BVGEF 121 AND METHODS FOR THE PRODUCTION OF BOVINE AND HUMAN  
; VASCULAR ENDOTHELIAL CELL GROWTH FACTORS, BVGEF120 AND BVGEF121  
; NUMBER OF SEQUENCES: 40  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/559,041  
; FILING DATE: 27-JUL-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 450,883  
; FILING DATE: 14-DEC-1989  
; APPLICATION NUMBER: 387,545  
; FILING DATE: 27-JUL-1989  
; SEQ ID NO:4:  
; LENGTH: 88  
5219739-4

Query Match 37.0%; Score 14.8; DB 7; Length 88;  
Best Local Similarity 73.1%; Pred. No. 6.2e+02;  
Matches 19; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 gcatgtcagcagtagatgagctacaag 26  
||| ||||| ||||| ||| |||||  
DB 29 GCTGTAGACGTCCATGAACTTCAAG 4

RESULT 10

US-08-061-889-11/c  
; Sequence 11, Application US/0806189  
; Patent No. 5654138  
; GENERAL INFORMATION:

; APPLICANT: LERMAN, MICHAEL I., LATIF,  
; APPLICANT: FARIDA, SZBAR, BERTON, LINEHAN, MARSTON  
; TITLE OF INVENTION: VON HIPPEL-LINDAU (VHL)  
; TITLE OF INVENTION: DISEASE GENE AND CORRESPONDING CDNA AND  
; TITLE OF INVENTION: METHODS FOR DETECTING CARRIERS OF THE VHL  
; TITLE OF INVENTION: DISEASE GENE  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/061,889  
FILING DATE: 14-MAY-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD W. BORK

REGISTRATION NUMBER: 36,459  
REFERENCE/DOCKET NUMBER: 2026-4078

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800

TELEFAX: (212) 751-6849

TELEX: 421792

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 22 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-061-889-11

Query Match 36.5%; Score 14.6; DB 1; Length 22;  
Best Local Similarity 81.0%; Pred. No. 5.5e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 18 agctacaagcggtgattggg 38  
||||| ||||| ||| |||||  
DB 21 AGCTACAAGCATCACTGTGG 1

RESULT 11

US-08-462-611-11/c

; Sequence 11, Application US/08462611

; Patent No. 5759790

; GENERAL INFORMATION:

; APPLICANT: LERMAN, MICHAEL I., LATIF,  
; APPLICANT: FARIDA, SZBAR, BERTON, LINEHAN, MARSTON

; TITLE OF INVENTION: VON HIPPEL-LINDAU (VHL)

; TITLE OF INVENTION: DISEASE GENE AND CORRESPONDING CDNA AND

; TITLE OF INVENTION: METHODS FOR DETECTING CARRIERS OF THE VHL

; TITLE OF INVENTION: DISEASE GENE

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN & FINNEGAN

; STREET: 345 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10154

; COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/462,611  
FILING DATE: 05-JUNE-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/061,889  
FILING DATE: 14-MAY-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD W. BORK  
REGISTRATION NUMBER: 36,459  
REFERENCE/DOCKET NUMBER: 2026-40780S1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-462-611-11

Query Match 36.5%; Score 14.6; DB 2; Length 22;  
Best Local Similarity 81.0%; Pred. No.5.5e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 18 agctacaagcgatgattggg 38  
|||||  
Db 21 AGCTACAAGCATCATCTGTG 1

RESULT 12  
PCT-US94-05378-11/C  
Sequence 11, Application PC/TUS9405378  
GENERAL INFORMATION:  
APPLICANT: THE GOVERNMENT OF THE UNITED  
STATES OF AMERICA AS REPRESENTED BY THE  
APPLICANT: SECRETARY, DEPARTMENT OF HEALTH AND HUMAN  
APPLICANT: SERVICES  
TITLE OF INVENTION: VON HIPPEL-LINDAU (VHL)  
TITLE OF INVENTION: DISEASE GENE AND CORRESPONDING CDNA AND  
TITLE OF INVENTION: METHODS FOR DETECTING CARRIERS OF THE VHL  
TITLE OF INVENTION: DISEASE GENE  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/05378  
FILING DATE: 12-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/061,889  
FILING DATE: 14-MAY-1993  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD W. BORK  
REGISTRATION NUMBER: 36,459

REFERENCE/DOCKET NUMBER: 2026-4078PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
PCT-US94-05378-11

Query Match 36.5%; Score 14.6; DB 6; Length 22;  
Best Local Similarity 81.0%; Pred. No.5.5e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 18 agctacaagcgatgattggg 38  
|||||  
Db 21 AGCTACAAGCATCATCTGTG 1

RESULT 13  
US-08-538-875-70/C  
Sequence 70, Application US/08538875  
Patent No. 5773582  
GENERAL INFORMATION:  
APPLICANT: Shin, Hang-Cheol  
APPLICANT: Shin, Nam-Kyu  
APPLICANT: Lee, Inkyung  
TITLE OF INVENTION: TUMOR NECROSIS FACTOR MUTAINS  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Shin, Hang-Cheol  
STREET: Jukong Gocheung Apt. 1014-806, Haan-dong  
CITY: Kwangmyung-shi  
STATE: Kyungki-do  
COUNTRY: Republic of Korea  
ZIP: 423-060  
ADDRESSEE: Shin, Nam-Kyu  
STREET: #181-404 Sadang-4-dong, Dongjak-ku  
CITY: Seoul  
STATE:  
COUNTRY: Republic of Korea  
ZIP: 156-094  
ADDRESSEE: Lee, Inkyung  
STREET: 11/2, #302-39 Juan-4-dong, Nam-ku  
CITY: Incheon  
STATE:  
COUNTRY: Republic of Korea  
ZIP: 402-204  
ADDRESSEE: Kang, Sungzong  
STREET: #84-4 Daesin-dong, Seodaemun-ku  
CITY: Seoul  
STATE:  
COUNTRY: Republic of Korea  
ZIP: 120-160  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette 3.5inch 2.0mb storage  
COMPUTER: IBM PC/AT  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/538,875  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/193,336  
FILING DATE:  
APPLICATION NUMBER: KR 93-1751  
FILING DATE: 9-FEB-1993

```

1  ATTORNEY/AGENT INFORMATION:
2  NAME:
3  REGISTRATION NUMBER:
4  REFERENCE/DOCKET NUMBER:
5  TELECOMMUNICATION INFORMATION:
6  TELEPHONE:
7  TELEFAX:
8  INFORMATION FOR SEQ ID NO: 70:
9  SEQUENCE CHARACTERISTICS:
10 LENGTH: 33 bases
11 TYPE: nucleic acid
12 STRANDEDNESS: single
13 TOPOLOGY: linear
14 MOLECULE TYPE: primer DNA
15 US-08-538-875-70
16
17 Query Match 36.5%; Score 14.6; DB 2; Length 33;
18 Best Local Similarity 69.0%; Pred. NO. 6e+02;
19 Matches 20; Conservative 0; Mismatches 9; Indels 0; Gaps 0.
20
21 QY 11 gtatgtgacctacaagcgatgttgga 39
22 ||||||| ||| ||| |||||
23 Db 33 GTAGATGAGCAAGCCCTCTATTGGCA 5
24
25 RESULT 14
26 US-08-600-783-8/C
27 Sequence 8, Application US/08600783
28 Patent No. 5962267
29
30 GENERAL INFORMATION:
31 APPLICANT: SHIN, Hang Cheol
32 APPLICANT: CHANG, Seung Gu
33 APPLICANT: KIM, Dae Young
34 APPLICANT: KIM, Chong Suh1
35 TITLE OF INVENTION: Proinsulin Derivative and Process
36 TITLE OF INVENTION: for Producing Human Insulin
37 NUMBER OF SEQUENCES: 36
38
39 CORRESPONDENCE ADDRESS:
40 ADDRESSEE: SHIN, Hang Cheol
41 STREET: Saangma-Hanshin Apt. 102-1206,
42 STREET: #245 Cholsan-dong
43 CITY: Kwangmyung-sh1
44
45 STATE: Kyungki-do
46 COUNTRY: Republic of Korea
47 ZIP: 423-030
48 ADDRESSEE: CHANG, Seung Gu
49 STREET: Hyundai Apt. 71-203, Apkujong-dong,
50 STREET: Kangnam-ku
51 CITY: Seoul
52
53 STATE: Seoul
54 COUNTRY: Republic of Korea
55 ZIP: 135-110
56 ADDRESSEE: KIM, Dae Young
57 STREET: Sosa Jukong Apt. 108-202, Sosa Bon-dong,
58 STREET: Sosa-ku
59 CITY: Bucheon-sh1
60
61 STATE: Kyungki-do
62 COUNTRY: Republic of Korea
63 ZIP: 422-230
64 ADDRESSEE: KIM, Chong Suh1
65 STREET: Garden Heights Apt. 202-801, #100,
66 STREET: Hwangkeum-dong, Soosung-ku
67 CITY: Taegu
68
69 STATE: Taegu
70 COUNTRY: Republic of Korea
71 ZIP: 706-040
72
73 COMPUTER READABLE FORM:
74 MEDIUM TYPE: Floppy Disk, 3.5 inch, 1.44MB storage
75 COMPUTER: IBM PC/AT
76 OPERATING SYSTEM: MS-DOS
77 SOFTWARE: word Perfect 5.1
78
79 CURRENT APPLICATION DATA:

```

```

1 APPLICATION NUMBER: US/08/600.783
2 FILING DATE:
3 CLASSIFICATION: 514
4 PRIOR APPLICATION DATA:
5 APPLICATION NUMBER: KR 95-2751
6 FILING DATE: 15-FEB-1995
7 ATTORNEY/AGENT INFORMATION:
8 NAME: Shahan Islam
9 REGISTRATION NUMBER: 32.507
10 TELECOMMUNICATION INFORMATION:
11 TELEPHONE: (212) 278-1000
12 TELEFAX: (212) 953-7249
13 INFORMATION FOR SEQ ID NO: 8:
14 SEQUENCE CHARACTERISTICS:
15 LENGTH: 33 bases
16 TYPE: nucleic acid
17 STRANDEDNESS: single
18 TOPOLOGY: linear
19 MOLECULE TYPE: primer DNA
20 ANTI-SENSE: no
21 US-08-600-783-8
22
23 Query Match 36.5%; Score 14.6; DB 4; Length 33;
24 Best Local Similarity 69.0%; Pred. NO. 6e+02;
25 Matches 20; Conservative 0; Mismatches 9; Indels 0; Gaps 0.
26
27 QY 11 gtatgtagctcaagcgctgattggga 39
28 ||||||| ||| ||| |||||
29
30 Db 33 gtatgtagagAACAGCCCTCTATTGGCA 5
31
32 RESULT 15
33 US-08-600-783-9
34 Sequence 9, Application US/08600783
35 Patent No. 5862267
36 GENERAL INFORMATION:
37 APPLICANT: SHIN, Hang Cheol
38 APPLICANT: CHANG, Seung Gu
39 APPLICANT: KIM, Dae Young
40 APPLICANT: KIM, Chong Suhll
41 TITLE OF INVENTION: Proinsulin Derivative and Process
42 TITLE OF INVENTION: for Producing Human Insulin
43 NUMBER OF SEQUENCES: 36
44 CORRESPONDENCE ADDRESS:
45 ADDRESSEE: SHIN, Hang Cheol
46 STREET: Ssangma-Hanshin Apt. 102-1206,
47 STREET: #245 Cholsan-dong
48 CITY: Kwangmyung-shi
49 STATE: Kyungki-do
50 COUNTRY: Republic of Korea
51 ZIP: 423-030
52 ADDRESSEE: CHANG, Seung Gu
53 STREET: Hyundai Apt. 71-203, Apjujong-dong,
54 STREET: Kangnam-ku
55 CITY: Seoul
56 STATE: Seoul
57 COUNTRY: Republic of Korea
58 ZIP: 135-110
59 ADDRESSEE: KIM, Dae Young
60 STREET: Sosa Jukong Apt. 108-202, Sosa Bon-dong,
61 STREET: Sosa-ku
62 CITY: Bucheon-shi
63 STATE: Kyungki-do
64 COUNTRY: Republic of Korea
65 ZIP: 422-230
66 ADDRESSEE: KIM, Chong Suhll
67 STREET: Garden Heights Apt. 202-801, #100,
68 STREET: Hwangkeun-dong, Soosung-ku
69 CITY: Taeju
70 STATE: Taegu
71 COUNTRY: Republic of Korea
72

```

ZIP: 706-040  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk, 3.5 inch, 1.44MB storage  
COMPUTER: IBM PC/AT  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/600,783  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: KR 95-2751  
FILING DATE: 15-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Shahar Islam  
REGISTRATION NUMBER: 32,507  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-1000  
TELEFAX: (212) 953-7249  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: primer DNA  
ANTI-SENSE: yes  
US-08-600-783-9

Query Match 36.58; Score 14.6; DB 4; Length 33;  
Best Local Similarity 69.08; Pred. No. 6e+02;  
Matches 20; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 11 gtagatgagctacaagcgctcatttgga 39  
||||||| ||| ||| |||||  
Db 1 GTAGATGAGCAGACAGCCCTCTATTGGCA 29

Search completed: June 4, 2000, 16:09:50  
Job time: 28069 sec

**THIS PAGE BLANK (USPTO)**